统计学原理

（第 3 版）

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

本书是经济管理类学科各专业的核心课程之一。主要介绍了该学科的性质、特点和基本概念，统计数据的收集与整理方法，相对指标、平均指标和变异指标的计算与分析方法，抽样推断的基本知识，相关与回归、指数和动态数列等常用统计分析方法。本书结构严谨，体系完整；语言简练，注重实用；强调计算机在统计中的应用。本书的前两版获得广泛好评，多次重印。

本书适合高职高专院校财会、管理、经济等专业师生，各类培训学校学员以及广大从业人员阅读。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

统计学原理 / 梁俊平，施燕萍著. —3 版. —北京：电子工业出版社，2009.8
21 世纪高职高专规划教材. 经济管理系列
ISBN 978-7-121-09078-3

Ⅰ. 统… Ⅱ. ①梁… ②施… Ⅲ. 统计学－高等学校：技术学校－教材 Ⅳ. C8

中国版本图书馆 CIP 数据核字（2009）第 100270 号

责任编辑：刘露明
印　　刷：
装　　订：
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：787×980　1/16　印张：15.25　字数：327 千字
印　　次：2009 年 8 月第 1 次印刷
定　　价：25.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

出 版 说 明

高职高专教育是我国高等教育的重要组成部分。近年来，高职高专教育有了很大的发展，为我国的现代化建设事业培养了大批急需的各类专门人才，为经济发展和社会进步起到了重要作用。

高职高专教育不同于其他传统形式的高等教育，它的根本任务是培养生产、建设、管理和服务第一线需要的，德、智、体、美全面发展的高级技术应用型专门人才，学生应在掌握必要的基础理论和专门知识的基础上，重点掌握从事本专业领域实际工作的基本知识和职业技能，因而对应这种形式的高等教育教材也应有它自己的体系和特色。

为了适应我国高职高专教育对教学改革和教材建设的需要，在国家教育部的指导下，电子工业出版社在全国范围内组织并成立了“全国高职高专教学研究与教材出版编委会”（以下简称“教学研究与教材出版编委会”），旨在研究高职高专的教学改革与教材建设，规划教材出版计划，以推动教育部策划的“21 世纪高职高专规划教材”的出版工作。“教学研究与教材出版编委会”的成员单位皆为教学改革成效较大、办学特色鲜明、办学实力强的普通高校、高等专科学校、高等职业学校、成人高等学校及本科院校主办的二级职业技术学院，而教材的编者和审定者则均来自于从事高职、高专和成人高等教育教学与研究工作第一线的优秀教师和专家。

为推动教育部策划的“21 世纪高职高专规划教材”的出版工作尽快实施，“教学研究与教材出版编委会”对高职高专教材的出版进行了规划。规划教材覆盖了计算机、通信、电子电气、财会和管理等专业的主要课程，主要面向课程包括基础课和专业主干课。这些教材全部按教育部制定的“高职高专教育基础课程教学基本要求”编写，适合于各类高等专科学校、高等职业学校、成人高等学校及本科院校主办的二级职业技术学院使用。

“教学研究与教材出版编委会”根据《教育部关于加强高职高专教育人才培养工作的意见》的文件精神，力求规划教材能够反映高职高专课程和教学内容体系改革方向，按照突出应用性、实践性的原则重组系列课程教材结构；力求教材能够反映当前教学的新内容，突出基础理论知识的应用和实践技能的培养。教材中的基础理论以应用为目的，以必需、够用为度，在专业课程教材的内容设计上加强了针对性和使用性；教材内容尽量体现新知识、新技术、新工艺、新方法，以利于学生综合素质的形成和科学思维方式和创新能力的培养。

编写高职高专教材是一个新课题，希望全国高职、高专和成人高等教育院校的师生在教学实践中积极提出意见与建议，并及时反馈给我们，以便我们对已出版的教材不断修订、完善，与大家共同探索我国高职高专教育的特点和发展道路，不断提高教材质量，完善教材体系，为社会奉献更多更新与高职高专教育配套的高质量的教材。

全国高职高专教学研究与教材出版编委会

E-mail: lmliu@phei.com.cn

第 3 版前言

《统计学原理》一书自第 1 版出版以来已经 7 年多了，在这段时间里，本书得到许多高等院校师生和其他读者的支持和好评。但是，在教材使用过程中也发现了一些问题，例如有些章节篇幅过大、内容复杂，不利于专科教学工作。特别是第 2 版的很多案例需要与时俱进。因此，我们认为有必要对第 2 版内容进行修订和增减。

第 3 版继承了本书第 1 版和第 2 版理论性和实践性相结合的特色，更加注重对实际经济现象的分析，同时借鉴很多国内外精品教材特色，在保持统计学完整结构体系的基础上删减了部分章节内容，修改了和当前实务不相符的相关案例；每章节开始都精选一个章首引例，有助于读者建立对本章内容的感性认识，同时，每章结束时都对重点内容进行总结，章节后面增加实践性环节练习，既有利于读者检验自学情况，又有利于读者参加该课程考试。

本书可以作为大专院校经济、管理以及财会专业课程教材，也可以作为企业统计人员培训和社会各界人员自学教材。

第 3 版教材由梁俊平和施燕萍担任主编，梁俊平对全书进行了统改与最后定稿。各章执笔人分别是：第 1 章由常州工学院梁俊平编写，第 2～4 章由常州工学院施燕萍和梁俊平编写，第 5 章由常州工学院马海涛在原来基础上改编，第 6 章由梁俊平在原来基础上进行了改编，第 7 章在原来基础上略作调整，第 8 章由常州工学院苏频编写。书稿的校对工作由梁俊平和马海涛共同完成。在本书再版之际，谨向帮助与支持本书编写与出版的同志表示衷心的感谢！

由于编者水平所限，书中难免会有疏漏或错误之处，恳请广大读者多提宝贵意见，以便我们进一步完善和修订。

编 者

第 2 版前言

统计学原理是经济、管理类学科各专业的基础课程，也是核心课程之一。这是一门定量研究社会经济现象的方法论学科。随着社会主义市场经济的迅猛发展和不断完善，越来越要求从事经济工作的人员能迅速提高管理水平和业务素质，以适应变化迅猛、竞争剧烈的市场经济的新形势。掌握统计学科中的定量分析方法，是对市场经济进行科学分析和正确决策的必要条件之一。统计学课程的一个重要特点和基本要求是理论必须密切结合实践。所以它又是一门提高我们分析和解决问题能力的应用性学科。

本书第 1 版出版后受到广大读者的喜爱，曾多次重印。这次在更新过程中，我们参考了国内外有关专著、教材，吸收了统计学教学和科研的新成果，充分考虑了统计改革的新经验和新成就。始终注重本学科的基本理论、基本知识的阐述及基本技能的培养和训练。其具有如下特点。

1．结构严谨，体系完整

本门课程包括统计数据的搜集、整理、描述、推论和分析等方法。具体到教材内容上，也正是按照这样的逻辑体系而组成的。大体上由以下几部分组成：

一是关于该学科的性质和特点（第 1 章），扼要介绍统计学的研究对象、特点和研究方法以及统计学中常用的一些基本概念；

二是统计数据的搜集与整理方法（第 2～3 章），介绍如何通过统计调查取得所需要的统计数据，并对所取得的统计数据进行必要的统计整理，以适合我们分析的需要；

三是描述统计方法（第 4 章），介绍相对指标、平均指标和变异指标的计算与分析方法；

四是推断统计方法（第 5 章），介绍有关抽样推断的基本知识；

五是经济与管理中常用的统计分析方法（第 6～8 章），重点介绍相关与回归，指数、动态数列等分析方法在经济领域中的应用。

2．语言简练，注重实用

本教材在写法上，力求概念准确、重点突出、简明扼要、深入浅出、通俗易懂，不着重于理论公式的推导和论证，而以数学理论的应用为主，以够用、会用为度。

3．强调计算机在统计中的应用

在第 1 版较为简单介绍 Excel 在统计中一些应用的基础上，此部分内容在新版中更全面、更详尽。之所以选择 Excel，是因为：

一是 Excel 很普遍，为大多数人所熟悉。而有些专业统计软件虽有很强的统计功能，但不如 Excel 普及。

二是对于高职高专层次所要求的统计学方法来说，Excel 的统计功能已能满足计算和分析的需要。

新版中，将在每章后面给出相应的、用 Excel 计算和分析的步骤或例题，提高学生运用计算机和统计方法分析、解决统计问题的能力。

4．更新了部分例题

对第 1 版中的例题与习题使用了更新的数据与内容，力求与现时的实际、热点相结合。为方便教学，本书配有教辅，内有多媒体教学课件、习题参考答案（可在电子工业出版社华信教育资源网 www.hxedu.com.cn 上下载）。

本教材由施燕萍担任主编，并对全书进行了统改与最后定稿。各章执笔人分别是：第 1～4 章由常州工学院施燕萍编写，第 5～6 章由太原工业经济学校武希礼编写，第 7 章由长治职业技术学院吕荣编写，第 8 章由常州工学院苏频编写。在本书再版之际，谨向帮助与支持本书编写与出版的同志表示衷心的感谢！

由于编者水平所限，书中有疏漏或错误之处，恳请同行和读者多提宝贵意见，以便我们进一步修改和完善。

编 者

2005 年 5 月

目　录

第 1 章　导论……1

1.1　统计的含义和作用……2
1.2　统计研究的对象、方法和过程……3
1.3　统计学中的一些基本概念……7
本章小结……11
复习思考题……11

第 2 章　统计数据的收集、整理和显示……14

2.1　统计数据的收集……15
2.2　统计数据的整理……24
2.3　统计数据的显示……36
本章小结……43
复习思考题……43

第 3 章　统计数据的描述……47

3.1　总量指标……48
3.2　相对指标……50
3.3　集中趋势的描述 ——平均指标……55
3.4　离散趋势的描述——变异指标……70
本章小结……75
复习思考题……75

第 4 章　抽样估计……80

4.1　抽样估计的一般问题……81
4.2　抽样误差……85
4.3　抽样估计的方法……92
4.4　抽样组织设计……100
本章小结……108
复习思考题……108

第 5 章　相关分析与回归分析……115

5.1　相关分析的意义和内容……115
5.2　相关关系的测定……118
5.3　一元线性回归分析……122
5.4　应用相关分析与回归分析应注意的问题……129
本章小结……130
复习思考题……130

第 6 章　时间数列分析……135

6.1　时间数列的编制……135
6.2　时间数列的水平指标……138
6.3　时间数列的速度指标……148
6.4　现象变动的趋势分析……156
本章小结……168
复习思考题……169

第 7 章　统计指数……174

7.1　统计指数及其种类……174
7.2　总指数的编制与应用……176
7.3　指数体系与因素分析……186
7.4　指数数列……196
本章小结……200
复习思考题……200

第 8 章　Excel 在统计分析中的运用……206

8.1　Excel 数据分析程序的安装 ……206
8.2　用 Excel 整理统计数据 ……208
8.3　用 Excel 计算描述统计量……212
8.4　用 Excel 进行区间估计 ……216
8.5　用 Excel 进行相关分析与回归分析……218
8.6　用 Excel 进行时间数列分析 ……222
8.7　用 Excel 进行指数分析……228
本章小结……230
复习思考题……230

附录　附表……232

参考文献……236

CHAPTER 1

第 1 章　导论

章首引例

2009 年 3 月 28 日，利用“Google”搜索引擎对因特网有关包含“统计”网页进行搜索，得到的结果是：包含“统计”这一词汇的所有网页有 282 000 000 项（其中包括 142 000 000 项中文网页），包含“金融危机”这一热门词汇的网页有 25 400 000 项（其中包括 7 890 000 项中文网页），前者是后者的 11 倍多。

在瑞典皇家设立的诺贝尔经济学奖项中，历届所有获此全球最高奖项殊荣的获奖者中有 2/3 以上的研究成果都与统计和定量分析有关。因此，著名经济学家萨缪尔森在经典教科书《经济学》（第 12 版）中特别提到，“在许多与经济学有关的学科中，统计学特别重要”。

美国杜邦公司的前CEO理查德森曾经指出：“现代公司的许多方面是根据统计数据进行科学决策的”。[①]

1981 年，首届国际《红楼梦》研讨会在美国召开，威斯康辛大学的陈炳藻教授独树一帜，宣读了一篇题为《从词汇上的统计论〈红楼梦〉作者的问题》的论文。他从字词出现频率入手，通过计算机统计分析，对《红楼梦》后 40 回系高鹗所著的流行看法提出质疑，认为整个 120 回合均出自曹雪芹之手。

本书专门阐述如何对现象（特别是经济现象）进行数据收集、数据整理以及数据分析的理论和方法。我们将在本章介绍统计的研究对象与基本研究方法。

本章学习目标

- ☑ 领会统计的含义、产生与发展、特点；
- ☑ 理解统计学的研究对象、研究过程与方法；
- ☑ 掌握统计总体、单位、标志、指标、指标体系等常用的基本概念。

① 转引自《马夸德特谈统计学家的重要性》，载《统计教育》，1994（3）。

1.1 统计的含义和作用

1.1.1 统计的含义

“统计”一词从字面上理解，“统计”是对大量事物的数量进行汇总和计数，即“统而计之”的意思。后来，从反映社会活动方面理解，“统计”一词延伸为用数字反映现象，表明统计还有用数字说明问题的作用。现代的“统计”一词，可以从三个不同方面理解，一般包括含义：统计工作、统计资料、统计学。

- 统计工作，指具体从事统计设计、资料收集、整理，并进行分析研究、预测和提供各种统计资料和统计咨询的实践活动的总称。
- 统计资料，指在统计工作过程中所获得的各种有关数字资料以及与之相联系的各种变动规律。它的表现形式为各种统计表、统计图、统计报告、统计年鉴及其他有关统计数字信息的载体等。其内容是反映社会经济现象的规模、水平、速度、结构、比例关系、变动规律等数字或文字资料。
- 统计学，是指阐述统计理论和方法的学科体系。从广义上讲，统计学包括自然科学和社会科学在内的统计科学理论的总和。而本书则专门阐述作为社会科学分支的统计学理论和方法，也可称为社会经济统计学，主要论述对社会经济现象如何进行统计设计、调查、整理和分析统计资料的理论和方法。

以上所述统计一词的三种含义是有着密切联系的。统计资料是统计工作的成果，统计学则是统计工作和统计资料的理论概括，而统计学形成的理论则对统计工作起着指导作用。统计工作一方面受统计理论指导，另一方面也检验着统计理论的正确与否，促进统计理论的不断发展。统计学与统计工作、统计资料之间的关系表明，理论来源于实践，又反过来为实践服务，被实践检验，体现着理论与实践辩证统一的关系。

1.1.2 统计的作用

1．统计是制定计划、实行宏观调控的基础

我国是社会主义国家，实行社会主义市场经济体制，对国民经济和社会发展仍须实行一定的计划管理，从基层单位直到整个社会都要制定计划，以指导经济建设和文化事业的发展。这就必须以正确的统计资料为依据，使计划工作建立在科学可靠的基础上。计划制定之后还要进行监督、检查，搞好经济预测，进行宏观调控，这些也要以统计为依据。

2．统计是制定政策的依据

各级党政领导机关在制定政策、方针时，都必须根据具体情况，从实际出发。如果离开了对实际情况的了解，想当然地制定政策，其后果是不可想象的。任何事物都是质与量

的辩证统一，从数量方面了解并掌握具体情况是制定政策时所必须注意的。正如毛泽东同志所说：“对情况和问题一定要注意到它们的数量方面，要有基本的数量分析……我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无‘数’，结果就不能不犯错误。”

3．统计是实行管理的手段

从社会发展而言，社会分工越来越细，生产技术不断进步，这就要求管理适应这种要求。社会主义社会是一个包含多部门、多层次的有机整体，各级领导要善于从错综复杂且多变的经济联系中，抓住主要环节，对瞬息万变的经济情况及时做出反应，这就需要各方面迅速而准确地提供信息。统计信息是各种信息的中心，是最重要的一种信息。因此，无论是宏观经济规划、管理、协调和平衡，还是微观管理、指挥调度和组织，一刻也离不开统计。各级统计机构必须及时而有效地提供统计信息，提供咨询，实施监督，参与监督，这是各级党政领导机关实施管理所不可或缺的。

4．统计是科学研究的工具

自然科学研究自然现象的主要方法是科学试验，可通过人为创造条件和控制，使其反复发生，进行观察、试验，其中要用到数理统计方法。作为社会科学的统计则是社会科学研究的主要工具。因为社会科学研究的对象是社会现象，对它不可能完全人为地加以控制和模拟，只能将某种社会经济现象的发生和发展忠实地“录”下来，然后进行统计分析研究，得出结论。社会科学的各学科的研究需要借助于“记录”事实的统计资料，进行分析比较研究，才能得到发展。

1.2　统计研究的对象、方法和过程

1.2.1　统计的研究对象

统计、统计学是从研究社会经济的数量方面开始的，随着统计实践的发展，统计方法的不断完善，统计学得以不断发展。这种理论与实践的关系，决定了统计学与统计工作的研究对象最终是一致的，都是社会经济现象总体的数量方面，即在质与量的辩证统一中研究大量社会经济现象总体的数量，反映社会经济现象发展变化规律在具体时间、地点和条件下的数量表现，揭示事物的本质、相互联系、变动规律性和发展趋势。但它们又有区别：统计工作是从实践中研究具体的社会经济现象发展规律；统计学则是从理论角度为统计工作提供数量研究和认识规律的科学方法，包括指导统计工作的原则、统计过程所应用的核算和分析方法，以及组织方法。其核心内容是数据的收集、整理、归纳、分析的原理和方法，这些方法论构成了统计学的科学体系。所以统计学就其性质来说，是研究如何收集数

据、分析数据，以便从中做出正确推断的方法论科学。

1.2.2 统计的特点

1．总体性

统计的研究对象不是个体现象的数量方面，而是许多个体现象构成的总体的数量方面。例如，进行城镇居民家计调查的目的不在于了解个别居民家庭的生活状况，而是要反映一个城市、一个地区、一个部门的居民收入水平、收入分配、消费水平消费结构等。当然统计研究又必须是从个体入手，如居民家计调查必须先从了解每个家庭的情况开始，然后经过分组、汇总，过渡到对城镇居民家庭总体特征的了解与认识。在认识和研究方法上遵循个体到总体的路径。

2．数量性

统计的研究对象是社会经济现象的数量方面，即用大量数字资料说明现象的规模、水平、现象间的数量关系，以及决定现象质量的数量界限等。统计研究对象的数量性，是区别于其他社会经济调查研究活动的根本特点。必须指出，统计对现象数量方面的认识，是与现象的质紧密结合在一起的，是以定性认识为基础，遵循定性—定量—定性的科学的认识规律。例如，要了解和研究工业增加值的数量及其变化，首先必须了解工业增加值的概念，然后才能根据这种认识去确定工业增加值的口径、范围和计算方法。

3．具体性

统计工作研究的总体数量是一个有具体时间、具体地点、具体条件限定的数值。例如，利润额 900 万元，在统计看来，它只是一个毫无意义的抽象数量。如果说 2008 年 9 月某企业利润额为 900 万元，这就是统计中所说的具体数量了。可见具体性就是指在时间、地点、条件三方面有着明确的规定性。

统计固然是研究具体的数量，但为了进行复杂的定量分析，还需要借助抽象的数学模型和数理统计方法，遵循一定的数学规则。因此说统计中对具体的数量研究需要密切联系抽象的数学方法。以抽象方法为手段，以具体数量为目的，体现了统计中具体和抽象的辩证关系。

1.2.3 统计的研究方法

统计是一门方法论科学，它提供对社会经济现象调查研究的理论、原则和方法。由于统计研究具有多阶段性，每个阶段都有许多特定的统计方法，归纳起来主要有以下几种基本方法。

1．大量观察法

大量观察法是指对所研究现象总体中的足够多数的个体单位进行观察，以期认识具有规律性的总体数量特征的方法，如统计报表、普查等。由于社会经济现象是在诸因素的综合作用下形成的，各单位的特征及表现有很大差别，不能只对少数单位进行调查。只有调查足够多的单位，才能从中认识客观现象的规律性。大量观察法主要用于统计调查阶段。

2．统计分组法

由于所研究现象本身的复杂性、差异性及多层次性，需要我们对所研究现象进行分组或分类研究，以期在同质的基础上探求不同组或类之间的差异性。统计分组在整个统计活动过程中都占有重要地位，在统计调查阶段可通过统计分组法来收集不同类的资料，并可使抽样调查的样本代表性得以提高（即分层抽样方式）；在统计整理阶段可以通过统计分组法使各种数据资料得到分门别类的加工处理和储存，并为编制分布数列提供基础；在统计分析阶段则可以通过统计分组法来划分现象类型、研究总体内在结构、比较不同类或组之间的差异（显著性检验）和分析不同变量之间的相关关系。

3．综合指标法

统计研究现象的数量方面的特征是通过统计综合指标来反映的。所谓综合指标，是指用来从总体上反映所研究现象数量特征和数量关系的范畴及其数值，常见的有总量指标、相对指标、平均指标和标志变异指标等。综合指标法用来描述统计学的核心内容。如何最真实客观地记录、描述和反映所研究现象的数量特征和数量关系，是统计指标理论研究的一大课题。

4．统计模型法

在以统计指标来反映所研究现象的数量特征的同时，我们还经常需要对相关现象之间的数量变动关系进行定量研究，以了解某一(些)现象数量变动与另一(些)现象数量变动之间的关系及变动的影响程度。在研究这种数量变动关系时，需要根据具体的研究对象和一定的假定条件，用合适的数学方程来进行模拟，这种方法就叫做统计模型法。

5．统计推断法

在统计认识活动中，我们所观察的往往只是所研究现象总体中的一部分单位，掌握的只是具有随机性的样本观察数据，而认识总体数量特征是统计研究的目的，这就需要我们根据概率论和样本分布理论，以一定的置信标准，由样本观测数据来推断总体数量特征。这种由样本来推断总体的方法就叫统计推断法。它可以用于总体数量特征的估计，也可以用于对总体某些假设的检验，所以在统计研究中应用很广泛，已成为现代统计学的基本方法。

以上五种方法相互联系，相辅相成，构成了一个完整的统计研究方法体系。

1.2.4 统计的研究过程

就一次完整的统计研究活动来讲，一般要经过统计设计、统计调查、统计整理和统计分析四个阶段。

1．统计设计阶段

统计设计是对统计活动各个方面和各个环节所做的通盘考虑和合理安排，如确定调查对象、设计指标体系、编制分类目录、制定调查、整理和分析方案等。优秀的统计设计是科学、有效地组织统计研究活动的前提。

2．统计调查阶段

统计调查就是根据一定的目的，通过科学的调查方法，收集社会经济现象的实际资料的活动，主要有统计调查方案的设计等。此为统计工作过程的第一阶段，实现统计设计的第一步，是认识客观经济现象的起点，也是统计整理和统计分析的基础。

3．统计整理阶段

统计整理是对调查来的大量统计资料加工整理、汇总，使之系统化、条理化，成为能够说明总体特征的综合资料的过程。它是统计工作过程的第二阶段，既是统计调查的继续，又是统计分析的前提，起着承前启后的作用。

4．统计分析阶段

统计分析是将加工整理好的统计资料加以分析研究，采用各种分析方法，计算各种分析指标，来揭示社会经济过程的本质及其发展变化的规律性，这是统计工作的第三阶段，通过第三阶段，对事物由感性认识上升到理性认识。

统计研究过程的四个阶段并不是孤立、截然分开的。它们是紧密联系的一个整体，其中各个环节常常是交叉进行的。例如，小规模的调查，常把调查和整理结合起来，在统计调查过程中就有对事物的初步分析，在整理和分析过程中仍须进一步调查（如图 1-1 所示）。

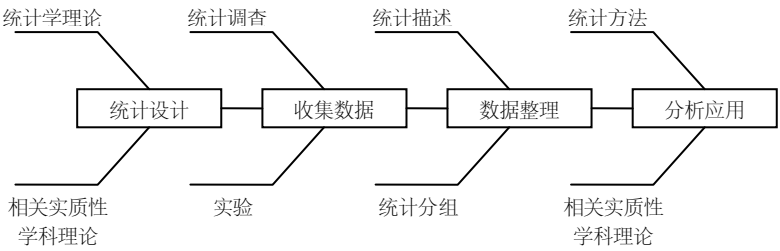


图 1-1　统计研究的过程描述

1.3 统计学中的一些基本概念

1.3.1 统计总体、总体单位和样本

1．统计总体

统计是研究客观现象总体的数量特征和数量关系。首先要对统计总体要有明确的认识。

统计总体就是根据一定目的确定的所要研究事物的全体。它是由客观存在的、具有某种共同性质的许多个别事物构成的整体（集合体），简称总体。例如，我们要研究全国乡镇工业企业发展情况，全国的乡镇工业企业就组成为一个总体。这些乡镇工业企业尽管资产规模、产品品种、技术力量、设备状况、经济效益等各不相同，但它们都是乡镇企业，都是工业生产单位，向社会提供工业产品或劳务服务，在这方面具有共同性，或称同质性，是统计总体赖以形成的客观基础。

总体具有以下三个特征。

（1）大量性

总体的大量性，是指总体应该由许多足够数量的同质性单位构成，而不能只有个别或少数单位。这是因为研究总体数量特征的目的是要揭露现象的规律性。而事物的规律性，特别是社会经济现象的规律性只有在大量现象的汇总综合中才能显示出来，个别单位的现象有很大的偶然性，而大量现象的总体则相对稳定，表现出共同性的倾向，这就是现象的必然性。

根据总体大量性的特点，总体可以分为有限总体和无限总体两类。无限总体是指包括的单位很多，以至无限的总体。例如，我们要研究海洋鱼类，海洋鱼类就是无限总体。又如在一条自动化加工的连续生产线上制造某种零件，这些零件可以被假定为永远不停地被制造出来，也属于无限总体。有限总体规模和范围相对较小，是由有限个单位事物构成的总体。例如某市工业企业组成的总体，高等学校在校学生组成的总体。社会经济统计中，大多数属于有限总体，可以采用全面调查，也可以用非全面调查。对无限总体，只能采用抽样调查来推断总体的情况。

（2）同质性

总体的同质性，是指构成总体的各个单位至少具有同一个性质，它是将总体各单位结合起来的基础也是总体的质的规定性。

同质性又是相对的，它是根据研究目的而定的，目的不同确定的总体就不同。例如研究企业职工的工资水平，全体企业职工构成总体，凡属企业职工都是同质的。但如果研究困难职工的工资水平，并确定困难线在年工资 5000 元以下，那么工资水平在这个界限上下的职工就属于不同质的了。

（3）变异性

总体各个单位除了具有某种或某些共同的性质以外，在其他方面则各不相同，具有质的差别和量的差别，这种差别称为变异。正因为变异是普遍存在的，才有必要进行统计研究，是统计的前提条件。总体中各个单位之间具有变异性的特点，这是由于各种因素错综复杂作用的结果，所以有必要采用统计方法加以研究，才能表明总体的数量特征。

2．总体单位

构成总体的每一个事物或基本单位就是总体单位，也称个体（元素）。原始资料最初就是从各个总体单位取得的，所以总体单位是各项统计数字最原始的承担者。例如，研究某个工业部门的生产情况时，该工业部门的所有工业企业可以作为一个总体，每个工业企业则是总体单位，将每个工业企业的某些数量特征加以登记汇总，就取得该工业部门的统计资料。

总体和总体单位是相对而言的，随着研究目的的不同，同一事物在不同情况下，可以作为总体，也可以作为总体单位。

3．样本

从某个总体中抽取部分单位所组成的整体，称为样本。样本中个体的数目称为样本容量。一般情况（不同领域标准不一）下，样本中个体的数目大于 30 称为大样本；样本中个体的数目小于或等于 30 称为小样本。在有些情况下，大样本和小样本所用的统计处理方法不一样。

1.3.2 标志和标志表现

1．标志

标志是反映总体各单位属性和特征的名称。每个总体单位从不同方面考察都具有许多属性和特征。例如，每个工人都具有性别、工种、文化程度、技术等级、年龄、工龄、工资等属性和特征，这些就是工人总体单位的标志。

2．标志表现

标志表现是总体各单位的某一标志之后所表明的属性或数值。标志分为品质标志和数量标志两种。凡是只能用文字表示的、表明单位属性方面的特征，如工人的性别、设备的种类、企业的经济类型等称为品质标志。性别标志具体表现为男女；设备的种类标志具体表现为车床、铣床、冲床等。凡是可用数值表示的、表明单位数量方面的特征，称为数量标志，如工人的工龄、工资，企业生产设备的能力，职工人数，产品产值等。工龄标志具体表现为不同的年数，某工人工龄是 15 年，则工龄是数量标志，15 年是该标志的数值表现。

不论品质标志还是数量标志，如果在一个总体的各单位具体表现都相同，就称之为不变标志。例如，在工人总体中，职业这一标志各单位表现都是工人，所以职业便是不变标志。在一个总体中，当一个标志在各单位的具体表现有可能不同时，这个标志便称为可变标志。例如，在工人总体中，各单位的工龄可能表现不同，所以工龄便是可变标志。可变标志的属性或特征的具体表现由一种状态变到另一种状态，统计上称之为变异，所以可变标志也称为变异标志。在一个总体中不变标志和变异标志各自发挥着重要的作用。一个总体至少要有一个不变标志，才能够使各单位结合成一个总体。

1.3.3　统计指标和指标体系

1．统计指标

统计指标是反映统计总体数量特征的概念和数值。例如，2007 年江苏省国内生产总值 25 741.15 亿元，其中，第二产业 14 306.40 亿元，占 55.58%；年末全省就业人员 4 618.14 万人；全省在岗职工年平均工资 27 374 元等。每一项统计指标都从规模、水平、速度、比例等方面反映江苏省国民经济和社会发展的数量特征。

由此可见，完整的统计指标是 6 项基本要素构成的：时间范围、空间、指标名称、核算方法、指标数值以及计量单位，其中指标名称和指标数值是核心两个要素。

- 指标的概念（名称），是统计所研究的社会经济现象的科学概念，表明社会经济现象的质的规定，反映某一社会现象内容所属的范围；
- 指标数值则是统计所研究现象的具体数量综合的结果，对某一社会经济现象总体特征从数量上加以说明。

统计指标必须包括时间状态、空间范围、计量方法等限定，不能随意变动，同时必须注意由于上述条件的变化而引起数值的可比性问题。总之，统计指标是统计指标名称及其指标数值的有机结合，是统计研究对象的具体化，也是统计对客观事物认识过程由质到量、质量结合的起点。

2．统计指标的分类

统计指标按其所反映的数量特点不同，可以分为数量指标和质量指标。凡是反映现象总规模、总水平和工作总量的统计指标称为数量指标，如人口总数、企业总数、职工总数、工资总额、国内生产总值、商品流转额、商品进出口总额等。这些指标不论是总体单位总量或是总体标志值总量都反映现象或过程的总规模和水平，所以数量指标也称为总量指标，用绝对数来表示。

凡是反映现象相对水平和工作质量的统计指标称为质量指标，如粮食平均亩产量、职工平均工资、人口密度、出生率、死亡率、工人出勤率等。质量指标是总量指标的派生指标，用相对数或平均数来表示，以反映现象之间的内在联系和对比关系。

3. 指标和标志的区别与联系

两者的主要联系是：

- 许多统计指标的数值是从总体单位的数量标志值汇总而来的。例如，我国钢产量(统计指标)是由我国每一个钢铁企业(总体单位)的钢产量(数量标志)的具体数值汇总而来的。所以，数量标志是统计指标的基础。
- 有些统计指标与数量标志之间在一定条件下存在着变换关系。由于研究目的不同，原来的统计总体可以变为总体单位，相应的统计指标就变成数量标志。例如，当把某企业作为统计总体时，其产量、职工人数、工资总额都是统计指标；当把该企业作为总体单位时，其产量、职工人数就成了数量标志。

两者的主要区别是：

- 说明的对象不同。指标是说明总体特征的，而标志则是说明总体单位特征的。
- 表示方法不同。标志有不能用数值表示的品质标志与能用数值表示的数量标志两种，而指标都是用数值来表示的，没有不能用数值表示的指标。

小提示：品质标志与质量指标、数量标志与数量指标有对应关系吗?

4. 指标体系

单个统计指标只反映总体某一个数量特征，说明现象某一侧面情况。但是客观现象是错综复杂、具有多方面联系的。要反映客观现象的全貌、描述现象发展的全过程，只靠单个统计指标是不够的，应该设立统计指标体系。

统计指标体系是由一系列相互联系的统计指标所组成的有机整体，用以反映所研究现象各方面相互依存、相互制约的关系。例如，工业企业是在一定生产经营主体的组织下，由资本金、劳力、物资、技术、设备、生产、供应、销售等相互联系的整体活动。为了反映企业生产经营的全貌，就应设立产量、产值、品种、质量、职工人数、劳动生产率、工资总额、原材料、设备、财务成本等指标群，来组成工业企业统计指标体系，其中，产品产量、总产值、增加值、品种、质量指标又构成企业生产统计指标体系，而固定资金、流动资金、生产费用、产品成本、销售利润又构成企业财务指标体系等。

社会经济统计指标体系大体上可以分为两大类，即基本统计指标体系和专题统计指标体系。基本统计指标体系是由社会指标体系、经济指标体系、科技指标体系三个子系统构成的，每个子体系又可以设若干门类指标或指标群，以反映其基本情况和相互联系。专题统计指标体系是指针对社会经济专门问题而制定的专项指标体系，如经济效益指标体系、国际收支指标体系、小康生活水平指标体系等。

指标体系的设置不但是客观现象的反映，而且也是人们对客观认识的结果。随着客观形势的发展变化以及实践经验和理论研究的积累，指标体系也将不断更新、日臻完善。

本章小结

1. “统计”一词有三种含义：统计工作、统计资料和统计学。统计工作是对客观事物数量方面的调查研究。在社会经济领域，通过统计调查、资料的整理和分析，以综合指标的形式反映社会经济现象在一定时间、地点、条件下的数量特征和数量关系，揭示现象发展变化的规律，为社会和各级政府管理部门提供信息和咨询，对国民经济和社会发展进行监测。在当今信息时代，统计信息是社会信息的主体。统计不只是认识社会的有力武器，而且也是科学决策和管理的重要工具。

2. 统计研究对象。在质与量的辩证统一中研究大量社会经济现象总体的数量方面规律。具有总体性、数量性和具体性的特征。

3. 统计研究的基本方法。从统计工作过程看，统计在各个不同阶段有着不同的工作内容和要求，相应地，就需要运用各种不同的统计研究方法，其中最基本方法有大量观察法、统计分组法、综合指标法、统计模型法和统计推断法。

4. 统计研究过程。一个完整的统计研究过程一般包括统计设计、统计调查、统计整理及统计分析四个阶段过程。

5. 统计学中最基本的概念是统计总体和总体单位、样本、标志和标志值、指标和指标体系。对这些概念必须正确理解，根据具体的研究任务和对象加以确定。

复习思考题

一、概念题

统计　统计学　大量观察法　综合指标法　统计推断法　总体　总体单位　标志　数量标志　品质标志　统计指标　指标体系

二、简答题

（1）怎样理解统计的含义、它们之间有什么关系?

（2）统计研究对象是什么? 它有哪些特点?

（3）统计工作过程分为哪几个阶段?

（4）统计研究的基本方法包括哪些?

（5）什么是统计总体和总体单位? 试举实际例子说明。

（6）什么是指标? 指标和标志有什么区别和联系?

（7）什么是统计指标体系?

三、练习题

1. 判断题（把“√”或“×”填在题后的括号里）

（1）统计调查过程中采用的大量观察法，是指必须对研究对象的所有单位进行调查。（　　）

（2）社会经济统计所研究的领域必须是社会经济现象总体的数量方面。（　　）

（3）总体的同质性质是指总体中的各个单位在所有标志上都相同。（　　）

（4）对某市中小学教师的收入状况进行普查，该市中小学教师的工资水平是数量标志。（　　）

（5）品质标志说明总体单位的属性特征，质量指标反映现象的相对水平或工作质量，二者都不能用数值表示。（　　）

（6）由女学生组成的总体中，“性别”这个标志是不变标志，不变标志是构成总体的基本条件。（　　）

2. 单选题

（1）社会经济统计的主要特点是（　　）。

A. 社会性，总体性　　B. 抽象性，数量性

C. 具体性，总体性　　D. 数量性，总体性

（2）社会经济统计现象形成总体的必要条件是（　　）。

A. 总体单位的大量性　　B. 总体单位间的同质性

C. 总体单位间的差异性　　D. 总体单位的社会性

（3）对某城市工业企业未安装设备进行普查，总体单位是（　　）。

A. 工业企业全部未安装设备　　B. 工业企业每一台未安装设备

C. 每个工业企业的未安装设备　　D. 每一个工业企业

（4）已知某位教师的工龄是15年，这里的“工龄”是（　　）。

A. 变量　　B. 指标

C. 品质标志　　D. 数量标志

（5）以产品等级来反映某种产品的质量，则该产品的等级是（　　）。

A. 数量标志　　B. 数量指标

C. 品质指标　　D. 数量指标

（6）某地区商业企业数、商品销售总额是（　　）。

A. 连续变量　　B. 离散变量

C. 前者是连续变量，后者是离散变量　　D. 前者是离散变量，后者是连续变量

3．多选题

（1）统计学的研究方法主要有（　　）。

A. 实验设计　　B. 大量观察　　C. 统计分组　　D. 综合分析

E. 统计推断

（2）在全国人口普查中（　　）。

A. 全国人口总数是统计总体　　B. 每一户是总体单位

C. 人的年龄是变量　　D. 性别男或女是品质标志表现

E. 人口的平均寿命是统计指标

（3）下列变量属于连续变量的是（　　）。

A. 某高校学生总数　　B. 身高

C. 某企业固定资产总额　　D. 城乡居民储蓄存款余额

E. 某市发生的交通事故总数

（4）下列各项中，属于统计指标的有（　　）。

A. 某企业总资产额 2 000 万元　　B. 某学生统计学考试成绩 85 分

C. 某地区国民生产总值 150 亿元　　D. 某居民户的人均支出 500 元/月

D. 某市年末就业人口数

（5）下列统计指标中，属于质量指标的有（　　）。

A. 工资总额　　B. 单位产品成品

C. 出勤人数　　D. 人口密度

E. 合格品率

（6）有一统计报表如下：某市国有商业企业 650 家，职工总数 41 万人，上月的商品零售价总额 90 亿元，职工平均工资额为 1 500 元。其中，A 企业的零售额为 39 万元，职工人数 820 人，则报告中出现有（　　）。

A. 统计总体　　B. 总体单位　　C. 标志　　D. 指标

E. 变量

CHAPTER 2

第 2 章　统计数据的收集、整理和显示

章首引例

2008 年是极不平凡的一年。我国经济社会发展经受住了历史罕见的重大挑战和考验。在中国共产党领导下，全国各族人民迎难而上，奋力拼搏，战胜各种艰难险阻，改革开放和社会主义现代化建设取得新的重大成就。

——国民经济继续保持平稳较快增长。国内生产总值超过 30 万亿元，比上年增长 9%；物价总水平涨幅得到控制；财政收入 6.13 万亿元，增长 19.5%；粮食连续 5 年增产，总产量 52 850 万吨，创历史最高水平。

——改革开放深入推进。财税、金融、价格、行政管理等重点领域和关键环节的改革取得新突破。进出口贸易总额 2.56 万亿美元，增长 17.8%。实际利用外商直接投资 924 亿美元。

——社会事业加快发展，人民生活进一步改善。城镇新增就业 1 113 万人；城镇居民人均可支配收入 15 781 元，农村居民人均纯收入 4 761 元，实际增长 8.4%和 8%。

——全面夺取抗击特大自然灾害的重大胜利。成功举办北京奥运会、残奥会。圆满完成神舟七号载人航天飞行。

这些成就，标志着我们在中国特色社会主义道路上迈出新的坚实步伐，极大地增强了全国各族人民战胜困难的勇气和力量，必将激励我们在新的历史征程上继续奋勇前进。[①]

那么如何去收集现象资料，又如何去整理资料信息呢？这正是本章所要回答的。让我们带着这个问题，一起进入这章的学习吧！

本章学习目标

- ☑ 了解统计调查、统计整理和分配数列的意义与要求；
- ☑ 理解不同统计调查方式的特点与应用条件；
- ☑ 掌握统计调查方案的设计内容、分组方法、变量数列的编制方法；
- ☑ 熟悉统计表和统计图的设计要求。

① 摘自《2009 年中国政府工作报告》。



2.1 统计数据的收集

2.1.1 统计调查的意义与要求

统计调查就是数据资料的收集。它是根据统计研究预定的目的要求和任务，以大量观察，运用科学的调查方法，有计划、有组织地收集资料的过程。在统计工作整个过程的四个阶段中，统计设计是第一阶段，是对统计全过程的规划，是“务虚”的范畴。从实际工作来看，统计调查才是真正的第一阶段，是统计整理和统计分析的前提。统计调查工作的好坏不仅说明本身的工作质量，而且直接影响整个统计工作的质量。如果统计调查收集的资料不可靠，统计信息就会失真，统计分析很难做出正确的判断，用以指导实际工作，必然会产生误导，甚至造成严重的损失。

统计资料的收集方式有两种：一种是直接向调查对象收集反映调查单位的第一手统计资料，一般称为原始资料，或初始资料的收集；另一种是根据研究的目的，收集已经加工整理过的、能在一定程度上说明总体现象特征的资料，一般称为次级资料或第二手资料的收集。例如，从统计年鉴、会计报表、报刊杂志、广播电视上得到的资料。由于次级资料一般都是从原始资料过渡而来的，所以统计调查主要是指对原始资料的收集。



提供社会和经济数据的公开出版物

《中国统计年鉴》、《中国统计摘要》、《中国社会统计年鉴》、《中国人口统计年鉴》、《中国工业经济统计年鉴》、《中国农业统计年鉴》、《中国市场统计年鉴》及各省、市、地区的统计年鉴、《世界经济年鉴》、《世界银行发展报告》、国家统计局网站以及各省市地区政府统计局官方网站等。

利用搜索引擎，如百度、Google 等，在各国及各省、市、地区政府部门和企业的网站上，收集公布的电子版数据，但要注意数据的真实性和完整性。

统计调查的基本要求如下：

- 准确性，就是如实反映客观实际，这是保证统计资料质量的首要环节。如果统计资料不真实必将给统计各个阶段的工作带来不良影响。
- 及时性，就是时效性，即要求在统计调查方案中规定的时间内，尽快提供资料，如果资料收集不及时，就会贻误统计整理分析的时间，使统计失去应有的作用。
- 系统性，是指收集的资料有条理，合乎逻辑，不杂乱无章，便于汇总。

- 完整性，是指调查单位不重复、不遗漏，所列调查项目的资料收集齐全。若统计资料残缺不全，就不可能反映所研究对象的全貌和正确认识社会经济现象总体和特征，最终也就难以对社会经济现象的规律性做出明确的判断，甚至会得出错误的结论。
- 经济性，是指在满足一定准确度的要求下，能以最小的费用取得所需的资料，努力降低调查成本。

以上五个要求是相互结合、相互依存的，在每一项统计调查过程中，要根据实际情况，分清主次缓急，综合起来考虑。一般而言，以准为基础，力求准中求快，准快结合，以尽可能小的成本取得完整系统的资料。

2.1.2 统计调查的方案设计

在统计调查工作正式开始之前，必须事先设计一个切实可行、周密细致的调查方案。统计调查方案主要包括如下几项内容。

1. 调查目的

在调查方案中首先应明确本次调查的目的、任务、意义。调查目的是调查所要达到的具体目标，它所回答的是为什么调查，要解决什么问题，调查具有什么样的社会经济意义等。只有这样才能确定向谁调查、调查什么以及采用什么调查方法。确定调查目的时，应做到具体明确，突出中心，避免面面俱到，以提高统计调查的质量。

2. 调查对象和调查单位

调查对象，是指需要调查的现象的总体，是由许多性质相同的总体单位组成的。

调查单位，是构成调查对象的每一个个体单位，它是进行调查登记的标志表现的直接承担者。

确定调查对象，就是要确定调查的总体范围界限，避免因界限不清而导致调查登记的重复与遗漏，保证资料的准确性。确定调查单位，就是要确定向谁调查来取得所需的有关标志表现的具体资料。

明确调查单位，还必须把它与报告单位区别。报告单位亦称填报单位，它是负责向上报告调查内容、提交统计资料的单位。报告单位一般是在行政上、经济上具有一定独立性的单位，而调查单位可以是个人、企事业单位，也可以是物。根据不同的调查目的，调查单位与报告单位，有时是一致的，有时不一致。例如，进行工业企业普查，每个工业企业既是调查单位又是报告单位；进行工业企业职工基本状况普查，调查单位是工业企业的每一位职工，而报告单位是每个工业企业。

3. 调查项目

调查项目就是要确定调查的内容，即向调查单位调查什么。这应根据调查目的和调查

对象的性质、特点、变化来确定，它是由调查单位的一系列品质标志与数量标志构成的标志体系，即一份调查过程中应该获得答案的各种问题的清单。

4．调查表和调查问卷

将各个调查项目按照一定的顺序排列在一定的表格上，就构成了调查表。利用调查表，能够有条理地填写需要收集的资料，便于调查后对资料进行汇总整理。

调查表一般有两种形式，一种是一览表，是把许多调查单位填写在一张表上。在调查项目不多时，采用该类表式，较为简便，便于合计和核对数据，如表 2-1 所示。另一种是单一表，是每个调查单位填写一份，可容纳较多标志，又便于整理分类，一般用于调查项目较多的场合，如表 2-2 所示。统计调查要采用哪一种表式，是由调查目的、任务而定的。

表 2-1　身体发育状况调查表　　编号

检查序号	姓名	性别	出生年月日	年龄	身高	体重	胸围	呼吸差	肺活量	坐高

填表人　　填表日期

表 2-2　年末职工家庭就业人口调查表

家庭人口：　　人　就业人口：　　人

姓名	与户主关系	性别	年龄	工作单位	职业	职务职称

被调查户主姓名　　填表人　　填表日期

问卷调查是采用问答式的调查表格，它不要求被调查者签署真实的姓名，这样可以减轻被调查者的心理压力，回答问题自然坦白，从而保证调查结果符合客观实际情况。此类调查适合对意识形态或个人隐私方面问题的调查。

5．调查时间和调查期限

调查时间是指调查资料所属的时间，在统计调查中，如果所调查的是时期现象，就是明确规定调查资料所反映的起止日期。例如，调查 2008 年第一季度的石油产量，则调查时间是从 1 月 1 日起至 3 月 31 日。如果所要调查的是时点现象，调查时间就是规定的统计标准时点。如我国的第五次人口普查调查时点是 2000 年 11 月 1 日零时。调查期限是进行调查工作的时限，包括收集资料和报送资料的工作所需的时间，应尽可能缩短。如 2000 年人

口普查规定2000年11月1日至11月10日现场登记完毕，则调查期限为11月1日至10日共10天。

6．调查方法

统计调查方法是指收集调查对象原始资料的方法。即调查者向被调查者收集答案的方法。任何一种调查都必须采用一定的调查方法去收集原始资料，即使调查的组织形式相同，其调查方法也是可以不同的。常用的方法有直接观察法、报告法、采访法、问卷调查法、登记法、卫星遥感法等。

（1）直接观察法

它是由调查人员到现场对调查对象进行观察、清点和计量，以取得资料的一种方法。直接观察法能够保证所收集资料的准确性，如农产量抽样调查中，调查人员亲自参加抽选样本、实割实测、脱粒晾晒、过秤计量等工作以及时了解农作物产量。但这种调查方法需耗用大量的人力、物力、财力和较长的时间，而且无法用于对历史资料的收集。

（2）报告法

报告法是调查单位以各种原始记录和核算资料为基础，收集各种统计资料，按隶属关系逐级上报，提供统计资料的方法。我国现行的统计报表制度采用的就是这种方法。由于这种用方法取得的资料，有原始记录、统计台账或核算资料为依据，一般比较可靠和准确。

（3）采访法

采访法是根据被调查者的答复来收集统计资料，这种方法又可分为口头询问法和被调查者自填法和开调查会。这种方法的效果取决于被调查者对调查工作的态度、被调查者与调查者的合作关系等。如人口普查等常用这种方法。采用这种方法可获得较为准确的数据，但需要较多的人力和时间。

（4）问卷调查法

问卷调查法是根据统计调查的要求，将调查项目编制成调查问卷，交给调查单位，由被调查单位自愿按照要求书面回答，以取得统计资料的一种方法，这种方法目前在国内外被广泛采用。例如，广告业中的产品品牌问卷调查即是一种典型的问卷法。这种方法省时省力，但如果回收率较低或者反馈答案质量不高，就会影响数据的质量及结论的可靠性。

（5）登记法

登记法是由有关的组织机构发出通知，规定当事人在某事发生后到该机构进行登记，填写所需登记的材料。例如，人口的出生和死亡的统计及流动人口的统计就是采用规定当事人到公安机构登记的方法。

（6）卫星遥感法

卫星遥感法是利用卫星高度分辨辐射所提供的地面农作物绿度的资料，以估计农业产量的方法。例如，我国运用这种方法估计北方小麦产量已有十多年，取得了较好的效果。



小提示：你知道电话调查、网上调查吗？它们是怎样的调查方法？有什么特点？

7．确定调查的组织实施计划

调查组织工作主要包括：调查领导机构和调查人员的组织；调查步骤、调查的方式方法；调查前的各项准备工作，如调查教育工作、人员培训、文字印刷等；调查经费的来源和开支办法；调查资料的报送方法和公布调查结果的时间。对于规模较大而又缺少经验的统计调查，常常还需要进行试点调查，通过试点，检验和修正原调查方案，积累组织实施的经验。



改革开放以来，我国第二、第三产业发展很快，在国民经济中所占的比重越来越高、而传统农业为主的第一产业在国民经济中的比重逐年降低。由于对第二、第三产业，特别是一些新兴服务业发展状况的把握还不够全面，了解还不够深入，我国在 2003 年、2008 年分别进行了两次经济普查工作。开展经济普查的目的，就是为了全面掌握我国第二产业和第三产业的发展规模、结构和效益等情况，建立健全基本单位名录库及其数据库系统。这对研究制定国民经济和社会发展规划，优化经济结构，改进宏观调控，开拓新的就业渠道，提高人民生活水平，全面建设小康社会，具有重要意义；对改革统计调查体系，完善国民经济核算制度，健全统计监测和预警、预报系统，发挥重要作用。

2.1.3　统计调查的组织方式

社会经济现象是复杂的，调查对象是千差万别的，统计研究的任务是多种多样的，因此，为了要准确、及时、全面、经济地收集所需资料，就应根据不同的调查对象和调查目的，灵活地运用各种不同的调查方式方法。统计调查方式方法有多种，它们各有不同的特点和作用，可以从不同的角度对其进行分类，如图 2-1 所示。

1994 年全国统计工作会议提出，要以建立必要的周期性普查为基础，经常性的抽样调查为主体，同时辅之以重点调查、科学推算和少量的全面报表综合运用的统计调查方法体系。其中“建立必要的周期性普查为基础，经常性的抽样调查为主体”体现了与国际统计惯例接轨，而“辅之以重点调查、科学推算和少量的全面报表综合运用的统计调查方法体系”则考虑到了中国的特点，体现了中国的特色。

1．统计报表

统计报表是按照国家有关法规的规定，自上而下地统一布置，逐级提供基本统计数据

的一种调查方式。它要以一定的原始记录为基础，按照统一的表式、统一的指标、统一的报送时间和报送程序进行填报。在我国几十年的国家和地方政府部门政府统计工作中，统计报表是收集统计数据的一种重要方式，是统计数据的主要来源，已形成了一套完备的统计报表制度。

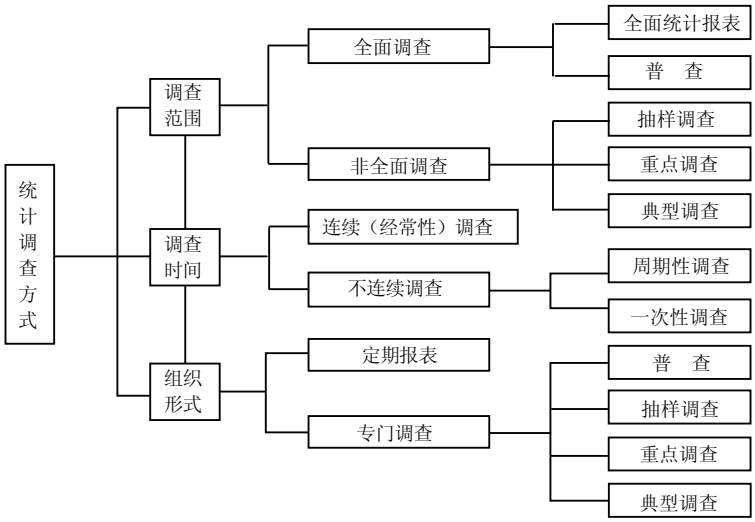


图 2-1　统计调查方式示意图

统计报表的类型较多，有全面与非全面、定期与临时之分，大部分及主要的统计报表都是全面的、定期的报表。定期的报表按报送时间不同可分为日报、月报、季报、年报等。按报送范围不同可分为国家、部门、地方统计报表。国家级的统计报表从国民经济整体出发，按国民经济的部门划分填报相应的指标，为整个国民经济的宏观管理服务；部门和地方统计报表按部门和地方的要求填报，主要为部门和地方管理服务。统计报表必须经过法定审批程序，凡依照国家法律规定审批和制发的统计报表，有关地区、部门和单位必须切实执行。

统计报表的资料，来源于基层单位原始记录。从原始记录到统计报表，中间还要经过统计台账和企业内部报表。因此，建立和健全原始记录、统计台账和企业内部的统计报表制度，是保证统计报表质量的基础。

与其他统计调查方式相比，统计报表有以下显著的特点：

- 统一性和时效性。为了保证统计资料的及时和完整，统计报表的指标内容、口径范围、计算方法和表格形式以及报送程序和报送时间都是国家统一规定的。
- 全面性。在报表的实施范围内各单位都必须全面地贯彻执行。这样，从基层单位到部门、地区直至中央国务院的填报和逐级汇总综合，可以得到国民经济全面的基本的统计资料。



- 相对的可靠性。统计报表的数据资料来源于基层单位的原始记录要建立、健全基层单位的原始记录制度就可以保证统计报表的可靠性。
- 统计报表定期上报，内容稳定，便于经常收集和积累资料，用来进行历史对比，较系统地分析社会经济现象发展变化的规律性。

虽然统计报表制度作为我国收集统计资料的一种主要调查组织形式，发挥了重要的作用，但也有局限性。例如，在逐级汇总上报过程中由于中间环节多，易受人为因素的干扰，较难保证数据的准确性；报表内容固定，不能反映出现的新情况等。所以有必要开辟多种调查渠道，采用各种不同的调查方法才能收集到所需的资料。

2．专门调查

专门调查是为了某些特定目的，研究某些专门问题而组织的调查。这种调查多属一次性调查，如普查、重点调查、典型调查和抽样调查。

（1）普查

普查是专门组织的一次性全面调查，如人口普查、工业设备普查、库存物资普查等。普查所收集的资料主要是表明某一现象在某一时点的情况，时间性要求很强。普查工作多在全国或较大范围内进行，需要动员大量人力、物力和财力。因此，一般用来收集那些不能够或不适宜用定期的全面统计报表收集的统计资料，以摸清国家重大的国情、国力的情况，为国家制定有关政策和措施提供依据。

普查的组织方式一般有两种：一是组织专门的普查机构，配备一定的普查人员，对调查单位进行直接的登记，如人口普查等；二是利用调查单位的原始记录和核算资料，颁发一定的调查表格，由填报单位进行填报，如库存物资普查等。

普查作为一种一次性的全面调查，对资料的准确性和时效性要求高。普查的面广量大，要求有更多的集中领导和统一行动。在组织普查工作中必须注意以下几点：

- 规定统一的标准时点。标准时点是指对被调查对象登记时所依据的统一时点。这个时点一经确定，所有调查资料都要反映这一时点上的状况，以避免收集资料时因情况变动而产生重复登记和遗漏的现象。例如，我国第一次经济普查的标准时点为 2004 年 12 月 31 日 24 时。
- 规定统一的普查期限。在普查范围内各调查单位或调查点尽可能同时进行普查，并尽可能在最短的期限内完成，以便在方法上和步调上保持一致性，以保证资料的准确性和时效性。例如，我国第一次经济普查，调查登记期限规定在 2005 年 1 月至 5 月完成。
- 规定普查的项目和指标。普查项目和指标一经规定，不准任意改变或增减，以免影响汇总综合，降低资料质量。同一种普查，每次项目和指标应力求一致，并按一定的周期进行。例如，国务院规定经济普查以后每 10 年进行两次，分别在逢 3、8 的年份实施。这样可更好地进行历次调查资料的对比分析，观察某种现象发展变化的情况。

快速普查是一种特殊的普查。就其进行的方式而言，属于第二种普查方式，即利用原始资料或核算资料由填报单位进行直接填报。它的目的主要是为了满足国家社会经济发展的迫切需要，普查任务布置和资料报送越过中间一切环节，普查资料直接报送到最高一级普查机构集中汇总。

相关链接

为了适应我国改革开放和建立社会主义市场经济体制的需要，从根本上改革统计调查体系，国务院于1994年7月20日批转国家统计局《关于建立国家普查制度，改革统计调查体系的请示》，明确了普查在统计调查体系中的基础地位，正式确立国家周期性普查制度。普查项目包括人口普查、农业普查、工业普查、第三产业普查和基本单位普查等5项。其中人口普查、第三产业普查、工业普查、农业普查每10年进行一次，分别在逢0、3、5、7的年份实施；基本单位普查每5年进行一次，在逢1、6的年份实施。在2003年，经国家统计局、国家发展和改革委员会、财政部共同研究，并报请国务院批准，决定调整国家普查项目和周期安排。

具体办法为：

① 将原定于2003年在全国进行的第三产业普查推迟，并与计划在2005年及2006年分别进行的工业普查、基本单位普查合并，同时将建筑业纳入普查内容，统称为经济普查，定于2004年在全国进行第一次经济普查。经济普查以企业事业组织、机关团体和个体工商户为对象，主要普查第二、第三产业的发展变化情况。该项普查以后每10年进行两次，分别在逢3、8的年份实施。

② 农业普查以从事第一产业活动的单位和农户为对象，主要普查农、林、牧、渔业的发展变化情况。因其涉及全国广大农村和农户，具体组织、填报和审核汇总的工作难度很大，需要单独实施。该项普查每10年进行一次，由逢7改为逢6的年份实施。

③ 人口普查以自然人为对象，主要普查全国人口和住房以及与之相关的重要事项。该项普查每10年进行一次，仍在逢0的年份实施。

（2）重点调查

重点调查是指只在调查对象中选择一部分重点单位进行调查，借以了解总体基本情况的一种非全面调查。重点单位是指其在总体中具有举足轻重地位的单位。这些单位数量虽少，但它们调查的标志值在总体标志值中占有绝大的比重。通过对这些单位的调查，就能取得反映总体基本情况的数量信息。例如，鞍钢、上钢、武钢、太钢、宝钢等几个钢铁企业，虽然在全国钢铁企业中只是少数，但它们的产量却占绝大比重。对这些重点企业进行调查，比全面调查要省时省力，且更能及时地了解全国钢铁生产的基本情况。



重点调查的应用条件是：总体中要有部分单位能较集中地反映所要研究的问题；调查的目的又是了解总体的基本情况。重点调查可以是专门组织的调查，也可以布置统计报表由重点单位填报。由于重点单位与一般单位差别很大，重点调查的结果不能用来推断总体的指标数。

（3）典型调查

典型调查就是根据调查的目的和要求，在对研究对象进行全面分析的基础上，有意识地选择部分有代表性的单位进行调查，以认识事物发展变化的规律性的非全面调查。这种调查与其他调查方法相比，具有这样的特点：

- 典型调查是从典型入手，逐步扩大到认识事物的一般性和普遍性，所以，典型调查可以取得代表性较高的资料。
- 典型调查选取的是少数单位，因此，调查可省时，省力，调查方法也较灵活，提高调查的时效性。
- 典型调查是深入细致的调查。通过对具有典型性和代表性的调查单位进行深入细致的调查研究，既可以收集到有关的准确的调查资料，又可以掌握具体、生动的情况，研究事物发展变化的过程和结果。同时可以抓住新的问题，了解新的情况，研究新生事物。

典型调查的作用在于：

- 可以研究新生事物或典型事例的情况，也可以作为其他统计调查的补充；
- 在一定条件下，可验证全国调查数字的真实性；
- 有时利用典型调查的资料来估算总体数字。

典型调查有两种选择方式：

- 如果在调查单位之间差异较少时，可选择一两个典型单位进行“解剖麻雀”式的调查；
- 如果作为调查对象的各单位之间差异较大时，可采用“划类选典”的办法，把总体分成若干类型，然后在每一类型中选择典型进行调查。

（4）抽样调查

抽样调查是指根据随机原则从调查总体中抽取部分单位为样本进行观察并根据其结果推断总体数量特征进行有一定可靠程度的推算的一种非全面调查的方法。

抽样调查的特点如下：

- 选取调查单位是本着随机的原则，排除主观因素的影响，这也是抽样调查的基本要求和进行推断的基础；
- 对总体的数量特征可以做出科学的估计；
- 抽样误差可以事先计算和控制。

与其他调查方式相比，抽样调查既能节省人力、物力、财力，又可以提高资料的时效性，而且能取得比较正确的全面统计资料，具有许多优点。因此，抽样调查是非全面调查中较完善和具有科学根据的方式，在市场经济条件下使用非常广泛。

有关抽样调查的理论和方法将在本书第 4 章中详述。

上述各种统计调查方法分别具有不同的特点、作用及局限性，在实践中，要注意各种统计调查方法的综合应用和选择，才能收集到所需的统计资料。

2.2 统计数据的整理

统计整理就是根据统计研究的目的和任务，对统计调查所收集到的各种数据进行科学分类和汇总，为统计分析提供系统化和条理化的信息资料的过程。统计整理是统计研究过程中一个十分重要的中间环节，起着承前启后的作用。通过整理，可以将说明个体的、局部情况的原始资料转化为反映总体的、全局情况的综合资料，是统计分析之前的必要步骤。

2.2.1 统计整理的程序

1．根据研究目的设计整理汇总方案

统计汇总方案的设计包括两方面：一是对于总体的处理方法。即对总体进行各种分组，达到对总体具体而深刻的了解，便于以后的分析研究；因此，汇总方案要确定统计分组与分组体系。二是确定用哪些统计指标来说明总体。即根据研究目的，设计一套汇总表，用以对调查项目进行汇总。

2．汇总前对统计数据资料的审核

在对统计数据整理之前，必须对原始数据进行严格的审核，主要检查数据的完整性与准确性。完整性，即检查被调查的系统和应调查的项目是否齐全、有无遗漏，是否按规定的份数、时间上报。准确性，即检查原始数据是否真实，可靠，计算是否有错误等。其检查方法有逻辑性检查和计算检查。逻辑性检查主要是从定性角度审核数据是否符合逻辑，内容是否合理，各项目与数字之间有无相互矛盾的现象。比如，性别为“女”的人所填的与户主的关系是父子，对于这种违背逻辑的项目应予以纠正。计算检查是检查表中的各项数据在计算方法上有无错误。例如，各分项数据之和是否等于相应的合计数，各结构比例之和是否等于 1 或 100%，出现在不同表格上的同一指标数值是否相同，等等。

3．对数据资料的分组与汇总

按照一定的组织形式和方法，根据调查资料的性质与特点，划分为若干组，并加总，计算出各组的单位数和合计数，计算出各组指标和综合指标的数值。分组和汇总是统计数据整理的中心工作。



4．编制统计表、绘制统计图

将整理结果用统计表和统计图的形式反映出来，它可清晰地、简明扼要地表述统计资料的内容。

5．统计资料的积累和保管

2.2.2　统计分组

1．统计分组的概念

统计分组是根据统计研究的目的和任务，按照一定的标志将统计总体划分为若干个组成部分的一种统计方法。例如，社会经济结构可按产业发生的时序，划分为第一产业、第二产业和第三产业；全国人口按性别可分为男性人口和女性人口。

通过统计分组，能够达到组间差别性、组内同质性的分组效果。即一方面将总体的各个单位划分为若干性质不同的组成部分，使不同组的单位之间具有明显的差别；另一方面又把性质相同的总体单位组合起来，使同组内各单位具有相对的同质性。换言之，统计分组相对总体、相对分组标志是“分”；而相对个体、相对其他标志则是“合”。统计分组实际上就是在统计总体内进行的一种定性分类，它能够将一个较大范围的同质总体划分为若干个性质不同的、范围较小的同质总体（即组）。

统计分组不仅是统计整理的基础，也是使认识深化的必要前提和手段，借助统计分组，可以区分不同性质的现象类型；研究现象内部的结构及结构的变化，揭示出现象的本质特征及发展规律；分析现象之间的依存关系。因此，统计分组在统计研究中占有重要地位，是统计研究的基本方法之一。

2．统计分组的要求

根据统计分组的定义可知，统计分组有三个要素：母项，须划分的总体；子项，划分以后的类（组）总体；分组标志，进行统计分组的标准和依据。

进行统计分组，在技术上有三个基本要求：**周延性**，要求分组以后各子项之和应等于母项；**互斥性**，组与组之间要相互排斥，不能重合；分组**标志的同一性**，每一次分组只能以一个标志为划分依据，不能同时采纳两个或两个以上的标志为划分依据。遵守以上要求，就能达到组内同质性、组间差别性的分组效果，反之，就可能会出现分组上的混淆和矛盾，这是我们在统计分组中必须注意的。

3．统计分组的种类

统计分组可以按照不同的标志进行分类。分组的标志是划分资料的标准和依据，分组的标志选择是否得当，关系到能否正确地反映总体数量特征及其变化规律。统计分组主要有如下几种。

（1）按分组标志的多少，可分为简单分组和复合分组

1）简单分组与平行分组体系

简单分组就是对研究现象按一个标志进行分组，它只能从某一方面说明和反映事物的分布状况和内部结构。例如，为了了解企业职工基本情况，可以选择年龄、工龄、文化程度等标志进行简单分组，具体情况如下：

按年龄分组	按文化程度分组
20岁及20岁以下	大专及大专以上
21～35岁	中专、技工
36～50岁	高中
51～55岁	小学
	识字不多
	文盲

对同一总体选择两个或两个以上的标志分别进行简单分组，就形成平行分组体系。如为了深入了解我国固定资产构成的基本情况，可以按照经济领域、物质生产部门、经济类型、经济用途、使用情况及所有权进行分组，这六个简单分组相互联系、相互补充便构成平行分组体系：

按经济领域分组	按经济用途分组
物质生产领域	生产性固定资产
非物质生产领域	非生产性固定资产
按物质生产部门分组	**按使用情况分组**
工业	在用固定资产
农业	租出固定资产
建筑业	未使用固定资产
运输邮电业	不需用固定资产
商业物资供销业等	封存固定资产
按经济类型分组	**按所有权分组**
国有经济	自用固定资产
集体经济	租入固定资产
个体经济	
其他经济类型	

2）复合分组与复合分组体系

许多场合，要用两个或两个以上标志分组，即先按一个标志分组，在此基础上再按第二个标志分小组，又再层叠地按第三个标志分成更小的组，这称为复合分组。复合分组形成复合分组体系。例如，固定资产投资项目，先按经济类型分组，再按投资规模分组，形

成如图 2-2 所示的复合分组。

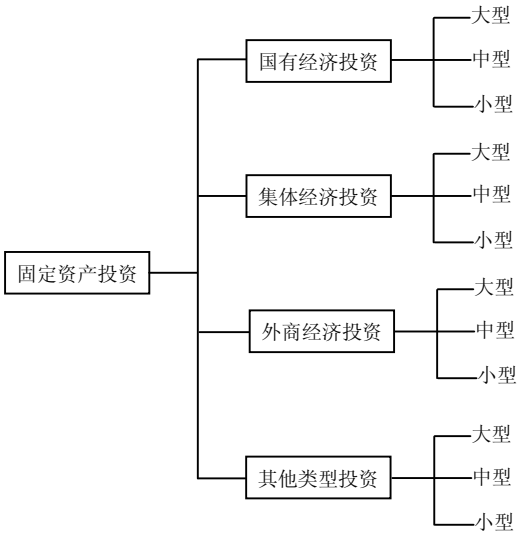


图 2-2　固定资产投资项目的复合分组图

复合分组和复合分组体系将多个标志层叠起来分组，能全面深入地说明问题。但当分组标志数目较多时，复合分组的组数将随分组标志的增加而成倍地增加，反而不易揭示出问题的实质。一般不宜采用太多的标志进行复合分组。

（2）按分组的标志的性质不同，分为品质分组（或称属性分组）和数量分组（或称变量分组）

品质分组就是按品质标志进行分组。一般地，对于类别数据，采用品质分组。例如，职工按性别分组，企业按经济类型分组等。数量分组就是按数量标志分组，数量标志的变异性体现在它不断变动自身的数量上，故也称为变量分组。例如，企业按产值、工人数分组。品质分组所形成的数列称为品质数列，变量分组所形成的称为变量数列。

（3）按分组的作用不同，分为类型分组、结构分组和分析分组

把复杂的现象总体，划分为若干个不同性质的部分，就是类型分组；在对总体分组的基础上计算出各组对总体的比重，借此研究总体各部分的结构，即结构分组。类型与结构分组往往紧密地联系在一起。

为研究现象之间依存关系而进行的统计分组为分析分组。分析分组的分组标志称为原因标志，与原因标志相对应的标志称为结果标志。原因标志不同，结果标志也会不同；同一原因标志由于分组的不同，结果标志也会不同。例如，工人的劳动生产率与产值之间、商品流通费用率与商品销售额之间的依存关系，都可以按分析分组法来研究它们之间的联系，如表 2-3 所示。

表 2-3　某地区部分商店按商品销售额分组的商品流通费用率表

商店按商品销售额分组（万元）	商店数（个）	商品流通费用率（%）
100 以下	10	9.8
100～300	12	8.7
300～500	11	7.5
500～700	9	6.5
700～900	8	5.8
900 以上	6	5.4

从表中可看出，随着商品销售规模的扩大，其商品流通费用率相应降低，两者表现出负依存关系。

4．统计分组的方法

统计分组的关键在于选择分组标志和确定各组的界限。

（1）正确选择分组标志

统计分组的核心问题就是如何正确地选择分组标志，这关系到能否确切地反映总体的特征，体现分组的科学性，实现统计研究的任务。列宁在论述选择分组标志的重要性时曾指出：“由于分类的方法不同，同一个材料可能得出完全相反的结论。”因此，为了正确地选择分组标志，必须注意以下几点：

- 要根据统计研究目的和任务选择分组标志，也就是应选择与统计研究任务密切相关的、最为符合统计研究目的的标志作为分组标志。同一研究对象研究目的不同，采用的分组标志也就不同。例如，为了研究某地区各类不同规模工业企业的生产经营状况时，可选择职工或生产能力作为分组标志。而研究目的在于确定该地区各种经济类型的工业企业的整个工业部门中所占的比重时，则可选择经济类型作为分组标志。
- 在总体若干个可供选择的标志中，要选择最能反映事物本质特征的标志作为分组标志。例如，研究居民的生活水平状况，可按城乡居民分组、按不同收入的居民分组，也可按居民的职业分组，还可以按脑力劳动者与体力劳动者分组等。在这些标志中，要注意选择主要的、起决定性的、能反映事物本质特征的标志作为分组标志。如上述按城乡分组和职业分组都是重要的分组。
- 要结合现象所处的具体历史条件和经济条件动态地选择分组标志。例如，企业按规模分组，而反映企业规模的标志很多，如职工人数、产品产量、产值、生产能力、固定资产价值等。选择哪一个作为分组标志，就必须结合企业所处的具体条件确定。在劳动密集型或技术不发达的条件下，宜选择职工人数作为分组标志；在技术密集型或技术装备比较先进的条件下，则宜采用生产能力或固定资产价值作为分组标志，这样才能确切地反映现象的本质特征。



社会经济现象是随着时间、地点和条件的变化而变化的，作为反映现象本质特征的分组标志，会随着现象性质特征的变化而失去其原来的重要意义。同一个分组标志适合于某一时间、地点、条件下的某现象，但不一定适合于另一时间、地点、条件下的该现象。因此，分组标志不能固定不变，即使是研究同类现象，也要视具体时间、地点、条件的不同，动态地加以选择，这样选择的分组标志才具有现实意义。

（2）正确确定各组的界限

分组标志确定后，就可以进一步在分组标志的变异范围内，具体划分各组的界限。分组标志按其形式，可分为品质标志和数量标志。统计总体可按品质标志分组，也可按数量标志分组。

- 按品质标志分组。就是选择反映事物属性差异的品质标志分组，并在品质标志变异的范围内，划分各组的性质界限，把总体区分为若干个性质不同的组成部分。按品质标志分组，一般来说概念、界限比较明确，区分容易，分组相对比较稳定。例如，人口按性别分为男、女两组；企业按经济类型分为国有、集体、个体、其他等组。但也有比较复杂的，具体表现为两种情况：一种是虽然按一个品质标志分组，但由于社会现象的复杂性，使得分组界限比较模糊，难以划分。例如，商品销售额按城乡分组，而处于城市与乡村结合部的地区其城乡特征不明显，区分起来就比较困难。另一种也是研究现象本身的复杂性所致，现象间出现交叉，往往须结合多个品质标志进行划分，其结果就形成一种复杂的类别。例如，国民经济按部门分类、人口按职业分类、产品按用途分类等。这种品质标志分组的方法，在统计上称之为分类法，应用十分广泛。对这些比较复杂的分组，事先须制定统一的划分标准或分类目录，以统一全国的分类口径。例如，我国统计部门制定的“工业部门分类目录”、“工业产品目录”、“主要商品目录”、“关于城乡划分标准的规定”等。
- 按数量标志分组。就是选择反映事物数量差异的数量标志分组，在数量标志变异的范围内，划分各组的数量界限，将总体区分为若干个性质不同的组成部分。例如，工人按工龄分组，企业按计划完成程度分组，商店按商品销售额分组，等等。数量标志的变异具体表现为许多不同的变量值，它们能准确地说明现象数量上的差异，但很难确切地反映现象性质上的差别。按数量标志分组的目的，就是要从各组的量的变化中反映出各组的质的特征，即通过现象数量上的差异和变化，根据变量值的变动特征，来划分性质不同的各组的界限，将社会经济现象的质的特征揭示出来。

按数量标志进行分组，其中涉及不少具体问题，如组距、组数和组限的确定，等距分组和异距分组等。由于按数量标志分组的结果，形成变量数列，这些问题将结合变量数列的编制给予说明。

2.2.3 次数分布与变量数列编制

1．次数分布

在按某一标志进行统计分组的基础上，将总体的所有单位按组归类排列，形成总体中各单位在各组间的分布，称为次数分布，或频数分布、分布数列。它是统计整理的一种重要形式，可用以研究总体各组分布状况，分布特征及总体的构成状况，还是进一步分析总体集中趋势和离散程度的基础资料。因此，编制分布数列，不仅是只反映统计整理结果的需要，也是进行统计分析的需要。其一般形式如表 2-4 所示。

表 2-4　次数分布的一般形式

分组标志	单位数（频数次数）	比率（或频率）（%）
	..	
合　计	总体单位数	100.00

根据分组特征的不同，分布数列可分为品质分布数列和变量分布数列两种。

（1）品质分布数列

品质分布数列是指按品质标志分组所形成的分布数列，简称品质数列。例如，根据我国第五次人口普查资料，大陆人口按性别标志分组，可编成品质数列，如表 2-5 所示。

表 2-5　第五次人口普查大陆人口的性别分布

性　别	人数（万人）	比率（%）
男　性	65 355	51.63
女　性	61 228	48.37
合　计	126 583	100.00
分组名称	频　数	频　率

品质数列属定类测定资料，如果分组标志选择得好、分组标准定得恰当，则事物的差异表现得就比较明确，总体中各组如何划分就容易解决。属性分布数列一般比较稳定，通常均能准确地反映总体分布特征。

（2）变量分布数列

变量分布数列是指按数量标志分组形成的分布数列，简称变量数列。例如，我国大陆人口按年龄分组可编制如下变量数列，如表 2-6 所示。

表 2-6　第五次人口普查大陆人口年龄分布

按年龄分组（岁）	人数（万人）	比率（%）
0～14	28 979	22.89
15～64	88 793	70.15
65 岁及 65 岁以上	8 811	6.96
合　计	126 583	100.00
各组变量值	频　数	频　率

表中，第 1 列是变量 x；第 2 列是各组单位数出现的次数 f，即频数，各组频数之和等于总体单位数；第 3 列是频率，是各组频数与总体单位总和之比，各组频率之和为 1。

$$\text{频率} = f / \sum f \qquad (2\text{-}1)$$

变量数列按照用以分组的变量的表现形式，可以分为单项式变量数列和组距式变量数列两种。

- 单项式变量数列指数列中每个组的变量值都只有一个，即一个变量值就代表一组，如表 2-7 所示。

表 2-7　某大学学生年看电影次数情况

按年看电影次数分组（次）	学生数（人）	比率（%）
0	200	4.44
1	500	11.11
2	1 500	33.33
3	2 000	44.44
4	300	6.66
合　计	4 500	100.00

- 组距式变量数列指按一定的变化范围或距离进行分组的变量数列，又称组距数列，如表 2-6 所示。

在组距式变量数列中，每一组的最大变量值称为该组的上限，最小变量值称为该组的下限。上限与下限之间的距离或差数就是该组的组距，即：组距=上限 – 下限。组距变量数列又有等距数列和不等距数列之分。如果各组组距都相等，称为等距数列，各组组距大小不等，称为不等距（或异距）数列。

2．变量数列的编制

（1）整理原始资料

变量数列的分组是按数量大小作为分组标准的。这样，就必须首先对原始资料按从小

到大的顺序排列，确定最大值和最小值，并计算全距。

根据抽样调查，某企业 50 名工人月平均收入资料如下（单位：元）：

1 250	1 320	1 230	1 100	1 180	1 580	1 210	1 460	1 170	1 080
1 050	1 100	1 070	1 370	1 200	1 680	1 250	1 360	1 270	1 420
1 180	1 030	870	1 150	1 410	1 170	1 230	1 260	1 380	1 510
1 010	860	780	1 130	1 140	1 190	1 260	1 350	930	1 420
1 080	1 010	1 050	1 250	1 160	830	1 380	1 310	1 270	880

上述资料比较零乱，不易直接看出其基本特征，若将这些数据按由小到大的顺序排列，可得到如下的阵列：

780	830	860	870	880	930	1 010	1 010	1 030	1 050
1 050	1 070	1 080	1 080	1 100	1 100	1 130	1 140	1 150	1 160
1 170	1 170	1 180	1 180	1 190	1 200	1 210	1 230	1 230	1 250
1 250	1 250	1 260	1 260	1 270	1 270	1 310	1 320	1 350	1 360
1 370	1 380	1 380	1 410	1 420	1 420	1 460	1 510	1 580	1 680

它反映出资料的某些特征：首先，说明月收入的波动幅度较大，其全距为 900 元。其次，说明多数工人的月收入在 1 000～1 400 元之间。通过整理，使我们大致对该资料的某些特征和基本状况有一个初步了解。

（2）确定变量数列的形式

对于离散型变量，因其所描述对象的数量特征，可以按一定的顺序一一列举数值，相邻两个变量之间不可能有小数。例如，高校的学生人数、机器台数、废品件数等。所以，对于这些变量，如果项数不多、变异幅度不大，可编制单项式变量数列；否则，应编制组距式变量数列。

对于连续型变量，因其所描述对象的数量特征，在一个区间内可以有无限多个数值，无法按一定次序一一列举，其变量值可以用小数表示。例如，粮食的亩产量、职工工资等。所以连续型变量不能编制单项式变量数列，而只能编制组距式变量数列。

（3）编制组距式变量数列应注意的问题

1）确定组距

组距的大小要适度，要能正确地反映总体的分布特征及其规律。组距与组数成反比例关系，组距越大，组数就越少；组距越小，组数就越多（组数 = 全距 ÷ 组距）。组数过少，容易把不同质的单位归在一个组内；组数过多，又容易把同质的单位分散在不同的组内，两者都不符合分组的要求。至于是采用等距分组还是采用不等距分组，要根据现象的特点、统计研究的目的及所收集到的资料分布是否均匀来确定。如果资料分布比较均匀，就可采用等距分组，否则应采用不等距分组。如上面所举工人月工资一例，则宜编制等距数列。



等距数列的组数、组距可以采用下列公式计算，作为参考。

$$K = 1 + 3.322 \lg n \qquad (2\text{-}2)$$

$$i = R / K \qquad (2\text{-}3)$$

式中，K 是组数；i 是组距；R 是全距；n 是数据个数。

2）确定组限

上限和下限统称为组限。确定组限的基本原则是：按这样的组限分组后，要能使性质相同的单位归入同一组内，使不同性质的单位按不同的组别划分。

对于离散型变量，其变量值都是整数，变量值之间有明显的界限，因而，组的上下限可用肯定性的数值表示，组限非常清楚。例如，工人按职工人数分组，其组限可表示为：

100 人以下
100～499 人
500～999 人
1 000 人以上

对于连续型变量，其变量值有小数，组限不能用肯定的数值表示，只能用前一组的上限与后一组的下限重叠的方法表示。例如，工厂按职工工资分组，可以表示如下：

900 元以下
900～1 100 元
1 100～1 300 元
1 300～1 500 元
1 500 元以上

一般原则是把达到上限值的单位划入下一组内。例如，当工资为 1 100 元时，该单位应属第三组而不是第二组。

在上述组限的表示方法中，数列的首末两组用“××以下”和“××以上”表示的叫开口组，首末两组上下限俱全的叫闭口组。在分组时是采用开口组还是闭口组，要根据现象的实际情况而定。

组中值是上限和下限之间的中点数值，它是代表各组标志值平均水平的数值。计算组中值的公式为：

$$\text{组中值}=(\text{上限}+\text{下限})/2 \qquad (2\text{-}4)$$

开口组的组距和组中值的确定，一般以其邻近组的组距为准，其计算公式为：

$$\text{缺下限开口组的组中值}=\text{上限}-(\text{邻组组距}/2) \qquad (2\text{-}5)$$

$$\text{缺上限开口组的组中值}=\text{下限}+(\text{邻组组距}/2) \qquad (2\text{-}6)$$

（4）频数分布表的具体编制

如前所举，该企业工人月平均收入的全距为 900 元，组距 = 全距 ÷ 组数=900 ÷ 5=180 元，可近似取 200。这里，组数取 5 组是根据研究的目的而定的。第 1 组为 900 元以下，表示最低收入；第 2 组为 900 ~ 1 100 元，表示较低收入；第 3 组为 1 100 ~ 1 300 元，表示收入为中等；第 4 组为 1 300 ~ 1500 元，表示收入较高；第 5 组为 1 500 元以上，表示收入高者，如表 2-8 所示。

表 2-8　50 名工人月平均收入频数分布表

按工人月平均收入分组 x（元）	频数 f	频率（%） $f / \sum f$	向上累计		向下累计	
			频数	频率（%）	频数	频率（%）
900 以下	5	10	5	10	50	100
900 ~ 1 100	9	18	14	28	45	90
1 100 ~ 1 300	22	44	36	72	36	72
1 300 ~ 1 500	11	22	47	94	14	28
1 500 以上	3	6	50	100	3	6
合　计	50	100	—	—	—	—

有时为了研究次数分布的状况，计算分析的需要，常需要计算累计次数或累计频率。计算累计次数或累计频率的方法有两种：一是向上累计，即把各组次数或频率由变量值小的组向变量值大的组的顺序逐项累计，各组的累计次数或累计频率表示小于该组变量值上限的次数或频率合计有多少。如上表中，第三组的向上累计次数和累计频率分别为 36 人和 72%，表示月平均收入低于 1 300 元的工人共有 36 人，占全部工人的 72%。另一种是向下累计，即把各组次数或频率由变量值大的组向变量值小的组的顺序逐项累计，各组的累计次数或累计频率表示大于该组变量值下限的次数或频率合计有多少。如上表中第二组的向下累计次数和累计频率分别为 14 人和 28%，表示月平均收入高于 1 300 元的工人共有 14 人，占全部工人的 28%。

通过对总体各单位分组而形成的变量数列，显示了各单位标志值在各组间的分布状况，从而使杂乱无章的原始数据显示出一定的规律性，从表中可看出，月平均收入在 1 100 ~ 1 500 元的工人占全部工人的 66%，而较低收入和高收入的工人所占比重较小，表现出近似“两头小，中间大”的钟形分布特征。

3．次数分布的主要类型

各种不同性质的现象有着各自特殊的次数分布。概括起来，主要有钟形、U 形、J 形分布三种。

（1）钟形分布

钟形分布的特征是“两头小，中间大”，即靠近中间的变量值分布的次数多，靠近两端的变量值分布的次数少。如果次数分配并不完全对称，则称为偏态分布，一般有左偏态和右偏态两种；如果次数分配完全对称，则称为对称分布或正态分布。正态分布是实际生活中最重要、最常见的分布，许多现象（如商品市场价格、农作物平均产量、零件公差等）统计总体的分布都趋于正态分布。对称分布和偏态分布如图 2-3 所示。

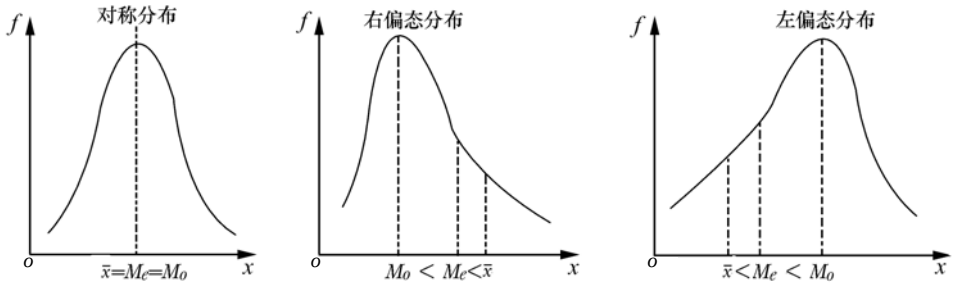


图 2-3　钟形分布图

（2）U形分布

U 形分布的特征是“两头大，中间小”，即靠近两端的变量值分布的次数多，靠近中间的变量值分布的次数少，如图 2-4 所示。人口死亡率、每日天空中云量的百分比等现象的分布都呈 U 形分布。

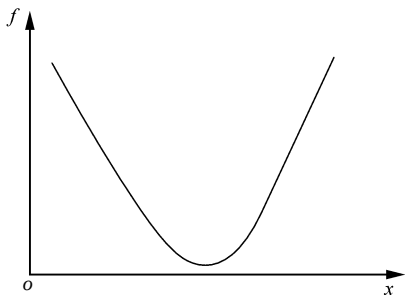


图 2-4　U 形分布图

（3）J形分布

J 形分布有正 J 形和反 J 形两种类型。正 J 形分布是次数随着变量值的增大而增多，如投资额按利润率大小分布；反 J 形分布是次数随着变量值的增大而减少，如人口总体按年龄大小分布，如图 2-5 所示。

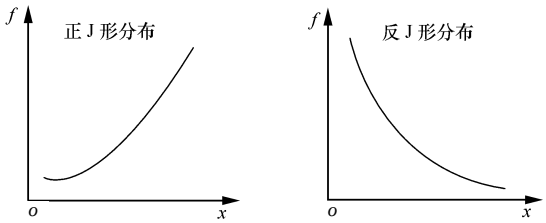


图 2-5　J 形分布图

2.3 统计数据的显示

2.3.1 统计表

统计调查所得来的原始资料，经过整理，得到说明社会现象及其发展过程的数据，把这些数据按一定的顺序排列在表格上，就形成了统计表，广义的统计表包括统计工作各个阶段中所用的一切表格。狭义的统计表专指分析表和容纳各种统计资料的表格，即通常所说的统计表。统计表优点在于，资料易于条理化，简单明了，可以节省大量文字叙述，便于比较分析，是统计分析的一种重要工具。

1．统计表的结构

统计表的结构，可以从形式和内容两个方面来认识。

（1）统计表的形式

统计表是由纵横交叉的线条组成的一种表格，表格包括总标题、横行标题、纵列标题和指标数值四个部分。总标题是统计表的名称，它扼要地说明该表的基本内容，并指明时间和范围，它置于统计表格的正上方；横行标题是横行的名称，一般放在表格的左方；纵列标题是纵栏的名称，一般放在表格的上方；横行标题和纵列标题共同说明填入表格中的统计数字所指的内容。指标数值是列在横行和纵列的交叉处，用来说明总体及其组成部分的数量特征，它是填写在统计表格的核心部分。

（2）统计表的内容

统计表由主词栏和宾词栏两个部分组成。主词栏是统计表所要说明的总体及其组成部分；宾词栏是统计表用来说明总体数量特征的各个统计指标。主词一般列在表的左方，宾词一般列在表的右方。必要时，主宾词可以变换位置或合并排列，如表 2-9 所示。

表 2-9　2004 年全国税收收入及其构成表

按税种分组	税收收入（亿元）	占总收入（%）
三项流转税 （国内增值税、消费税和营业税）	14 016	54.5

续表

按税种分组	税收收入（亿元）	占总收入（%）
三项所得税 （内资企业所得税、涉外企业所得税、个人所得税）	5 811	22.6
海关代征进口税收完成收入	3 706	14.4
其他税收收入	2 185	8.5
合　　计	25 718	100.0

主词栏　　宾词栏

此外，统计表还有补充资料、注解、资料来源、填表单位、填表人等。

2．统计表的种类

（1）按用途分类

广义的统计表可分为如下几类：

- 调查表，即在统计调查中用于登记调查项目的表格。
- 整理表或汇总表，即在统计整理汇总过程中使用的表格和用于表现统计汇总或整理结果的表格。
- 分析表，即用于统计分析的表格。这类表往往与整理表结合在一起，成为整理表的延续。

（2）按主词的结构分类

根据主词是否分组和分组的程度，分为简单表、分组表和组合表。

- 简单表：主词未经任何分组的统计表称为简单表，也称一览表。主词罗列各单位的名称。
- 分组表：主词只按一个标志进行分组形成的统计表，也称简单分组表。

简单分组表的应用十分广泛，对比简单表，它有如下作用：区分事物的类型，研究总体结构，分析现象之间的依存关系。

（3）复合表

主词按两个或两个以上标志进行分组的统计表，也称复合分组表。表 2-10 是复合表，表中国内生产总值分别按产业和国民经济行业分组。在复合分组表中设计横行标题时，应在第一次分组的各组组别下退一、二字填第二次分组的组别，这时第一次分组的组别就成为第二次分组的各组小计。若需再进行第三、第四次分组，均可按此类推。如表 2-10 所示，按产业进行第一次分组，分为第一、第二和第三产业。对第二产业又进行第二次分组，按行业分为工业和建筑业。

表 2-10　国内生产总值及其分组表

国内生产总值按产业和行业分组	国内生产总值（亿元）	比重（%）
第一产业		
第二产业		
工　　业		
建 筑 业		
第三产业		
交运仓储邮电通信业		
批发和零售贸易餐饮业		
合　　计		

（4）按宾词设计分类

统计表按宾词设计不同可分为宾词简单排列、分组平行排列和分组层叠排列等三种。

- 宾词简单排列是指宾词不加任何分组、按一定顺序排列在统计表上，如表 2-11 所示。

表 2-11　某地区不同经济类型工业企业主要经济指标

经济类型	企业数（个）	年平均职工人数（人）	工业增加值（万元）	年末固定资产净值（万元）
国有经济				
集体经济				
外商经济				
其他经济				
合　　计				

- 宾词分组平行排列是指宾词栏中各分组标志彼此分开，平行排列，如表 2-12 所示。

表 2-12　各地区社会商品零售总额　　单位：元

按地区分组	按商品性质和用途分组		按城乡分组		按经济类型分组			
	社会消费品零售总额	农业生产资料销售额	城镇	农村	国有	集体	个体	其他
北　　京								
天　　津								
河　　北								
山　　西								
内 蒙 古								
⋮								
合　　计								



- 宾词分组层叠排列是指统计指标同时有层次地按两个或两个以上标志分组，各种分组层叠在一起，宾词的栏数等于各种分组的组数连乘积。如表 2-13 所示，农村劳动力按产业分为 3 组，按性别分为 2 组，则复合分组设计的宾词栏数共有 3×2=6 栏（不包括合计栏）。

表 2-13　2005—2008 年农村劳动力的分布情况

年份	劳动力			产业								
	人　数			第一产业			第二产业			第三产业		
	合计	男	女	合计	男	女	合计	男	女	合计	男	女
2005												
2006												
2007												
2008												
总计												

统计表的主词分组与宾词分组是有区别的。主词分组的结果使总体分成许多组成部分，它们是需要用统计指标（宾词）来描述和表现的。宾词分组的结果并不增加统计总体的组成部分，仅仅是比较详细地描述总体已有的各个组成部分。由此可见，主词分组具有独立的意义，而宾词分组从属于主词的要求，是为描述主词的数量特征而考虑的。

3．统计表的编制应注意的问题

统计表设计总的要求是：简练、明确、实用、美观、便于比较。

（1）统计表形式设计注意事项

- 统计表应设计成由纵横交叉线条组成的长方形表格，长与宽之间保持适当的比例。
- 线条的绘制。表的上下端应以粗线绘制，表内纵横以细线绘制。表格的左右两端一般不画线，采用“开口式”。
- 合计栏的设置。统计表各纵列若需合计时，一般应将合计列在最后一行，各横行若需要合计时，可将合计列在最前一栏或最后一栏。
- 栏数的编号。如果栏数较多，应当按顺序编号，习惯上主词栏部分分别编以“甲、乙、丙、丁……”为序号，宾词栏编以（1），（2），（3），（4）等为序号。

（2）统计表内容设计注意事项

- 标题设计。统计表的总标题，横行、纵列标题应简明扼要，以简练而又准确的文字表述统计资料的内容、资料所属的空间和时间范围。
- 指标数值。表中数字应该填写整齐，对准位数。当数字因小可略而不计时，可写上“0”；当缺某项数字资料时，可用符号“…”表示；不应有数字时用符号“—”表示。

- 计量单位。统计表必须注明数字资料的计量单位。当全表只有一种计量单位时，可以把它写在表头的右上方。如果表中各栏的指标数值计量单位不同，可在横行标题后添一列计量单位。
- 注解或资料来源。为保证统计资料的科学性与严肃性，在统计表下，应注明资料来源，以便查考。必要时，还可在统计表下加上注解或说明。

2.3.2 统计图

运用统计图描述次数分布的类型特征，常用的图形有三种：直方图、折线图和曲线图。

现以某班 40 名学生统计学考试成绩资料（见表 2-14）为例，说明如何采用图示法来描述学生考试成绩的分布状况。

表 2-14 某班学生统计学考试成绩资料

按考分分组	次数 f_i	频率 $f_i / \sum f_i$	向上累计		向下累计	
			次数	频率（%）	次数	频率（%）
50～60	2	5.0	2	5.0	40	100.0
60～70	7	17.5	9	22.5	38	95.0
70～80	11	27.5	20	50.0	31	77.5
80～90	12	30.0	32	70.0	20	50.0
90～100	8	20.0	40	100.0	8	20.0
合　计	40	100.0	—	—	—	—

1. 直方图

即用直方形的宽度和高度来表示次数分布的图形。绘制直方图时，横轴表示各组组限，纵轴表示次数（一般标在左方）和比率（或频率，一般标在右方），若没有比率的直方图只保留左侧的次数。依据各组的组距的宽度与次数的高度绘成直方图。根据表 2-14 的资料绘制的直方图，如图 2-6 所示。

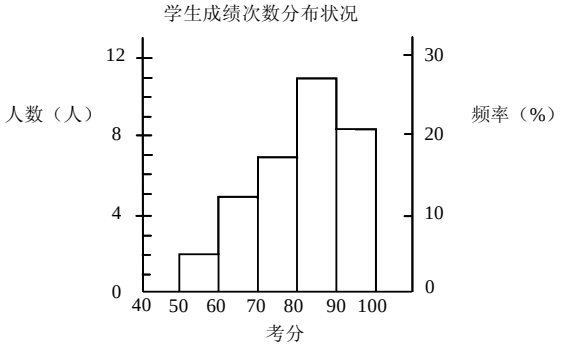


图 2-6 次数分布直方图

图 2-5 是依据等组距式变量数列绘制的直方图。对于不等组距式变量数列，则通常按次数密度（频数密度）绘制直方图以表示其分布。例如，40 名学生统计学成绩次数分布状况如表 2-15 所示。

表 2-15　某班学生统计学考试成绩次数分布表

按考分分组	组　距 d	次　数 f_i	频数密度=次数/组距
56～60	4	2	0.5
60～66	6	5	0.833
66～82	16	14	0.875
82～92	10	11	1.1
92～100	8	8	1.0
合　　计	—	40	—

对于不等距式变量数列，只有以频数密度绘制直方图，才能准确地反映客观实际情况。根据表 2-15 的资料，按频数密度绘制直方图，如图 2-7 所示。

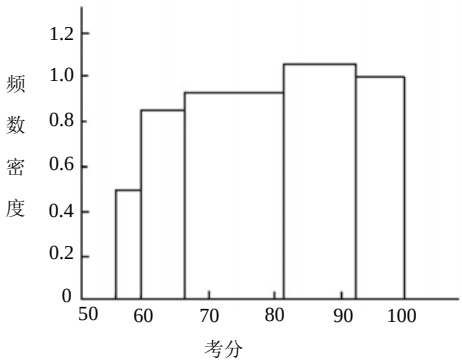


图 2-7　频数密度分布直方图

2．折线图

折线图可以在直方图的基础上，用折线将各组次数高度的坐标连接而成，也可以用组中值与次数求坐标点连接而成。图 2-8 是根据表 2-14 资料绘制的次数分布折线图。

3．曲线图

当变量数列的组数无限增多时，折线便近似地表现为一条平滑曲线。曲线图的绘制方法与折线图基本相同，只是在连接各组次数坐标点时应当用平滑曲线，而不用折线。对图 2-8 的折线平滑化，即得次数分布曲线图，如图 2-9 所示。

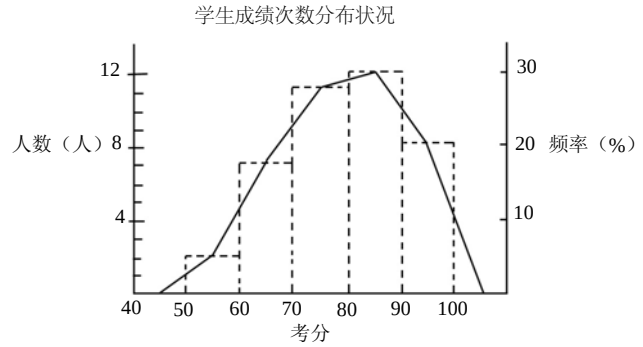


图 2-8　次数分布折线图

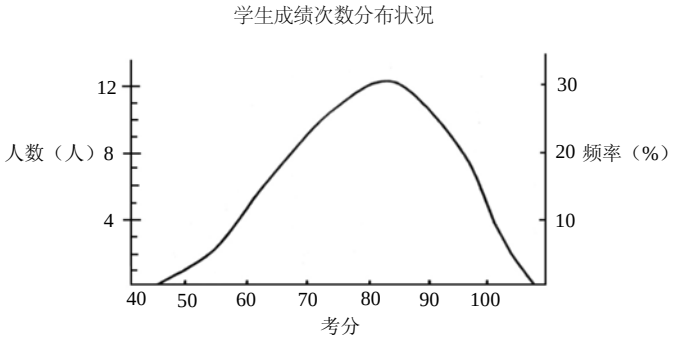


图 2-9　次数分布曲线图

4．条形图

条形图是用宽度相同的条形的高度或长短来表示数据变动的图形。它可以横置或纵置，纵置时也称为柱形图。在表示定类数据的分布时，用条形图的高度来表示各类别数据的频数或频率，如图 2-10 所示。

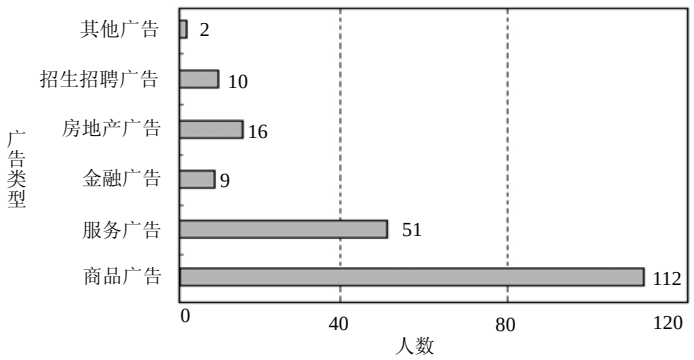


图 2-10　某城市居民关注不同类型广告的人数分布图

5．圆形图

圆形图也称饼图，是用圆形及圆内的扇形面积来表示数值大小的图形。圆形图主要用于表示总体中各组成部分所占的比例，对于研究结构性的问题很有用。这些扇形的中心角度是按各部分百分比占 360°的相应比例确定的。例如，关注服务广告的人数占总人数的百分比为 25.5%，那么其扇形的中心角度就应为 360°×25.5% = 91.8°，其余类推。根据图 2-10 数据绘制的圆形图如图 2-11 所示。

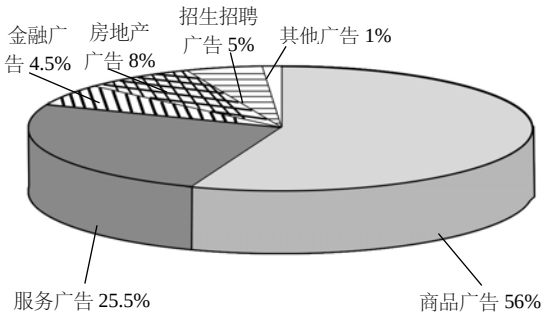


图 2-11　某城市居民关注不同类型广告的人数构成图

本章小结

1. 统计调查有普查、统计报表制度、重点调查、典型调查和抽样调查等多种组织方式，它们都有各自的特点和适用的条件，共同组成统计调查的方法体系。实际工作中，必须根据调查研究目的和具体条件选择调查方式。

2. 统计整理是连接统计调查和统计分析的重要阶段。统计资料整理的核心是分组。按品质标志分组的关键是界定各类型组的性质差异；按数量标志分组的关键是正确确定各组的数量界限，即确定组数和组距。必须使组与组之间的量的差异能够反映出局部的质的差异，通过分组形成的变量数列能够显示出数据的分布特征。

3. 现代统计调查的大量数据主要靠电子计算机进行整理，这是统计工作现代化的重要标志之一。无论是手工整理还是应用电子计算机进行数据处理，制作统计表和统计图必须注意规则，力求规范；否则，不仅统计资料不便于利用开发，而且还可能引起误解甚至得出错误的结论。

4. 统计表的编制应注意的问题主要分形式设计与内容设计两个方面。

复习思考题

一、概念题

普查　统计报表　抽样调查　重点调查　典型调查　次数分布　次数

二、简答题

（1）什么是统计设计，它有哪几种类型?
（2）统计设计的主要内容是什么?
（3）统计指标的特点是什么？它有哪些分类?
（4）设计统计指标体系时应该遵循哪些原则?
（5）设计统计表时应该注意哪些问题?
（6）什么是统计调查？它在统计工作中的地位如何?
（7）什么是调查对象、调查单位和填报单位？三者关系如何?
（8）统计调查有哪几种类型?
（9）普查、重点调查、典型调查和抽样调查各自的含义和特点是什么?
（10）什么是统计报表？它的作用是什么?
（11）设计统计表应该注意哪些问题?
（12）试举例说明什么叫全距、组距、组中值、上下限、频数、频率?

三、练习题

1．判断题（把“√”或“×”填在题后的括号里）

（1）为了研究某市的超市经营情况及存在的问题，需要对全市的超市进行全面调查。那么，该市所有的超市就是调查对象，每一个超市是调查单位。（　　）
（2）全面调查就是对调查对象的各方面都进行的调查。（　　）
（3）我国人口普查的总体单位和调查单位都是每一个人，而填报单位是户。（　　）
（4）我国第五次人口普查规定2000年11月1日零时为登记的标准时点，要求2000年11月10日以前完成普查登记。调查期限为10天。（　　）
（5）对一个企业来讲，整个企业统计工作的通盘安排是整体设计，而人力、物资、资金、生产等方面的设计就是单阶段设计。（　　）
（6）对连续大量生产的某种产品进行质量检查，最恰当的方法应该为抽样调查。（　　）

2．单选题

（1）统计工作过程大体上可以分为以下几个阶段（　　）。
A. 统计调查、统计整理和统计分析
B. 统计设计、统计调查和统计整理
C. 统计设计、统计调查、统计整理和统计汇总
D. 统计设计、统计调查、统计整理和统计分析
（2）重点调查、典型调查、抽样调查的根本区别是（　　）。
A. 调查对象包括的范围不同　　B. 调查的组织形式不同

C. 确定调查单位的方法不同　　D. 收集资料的方法不同

（3）统计设计按工作阶段可分为（　　）。

A. 整体设计和专项设计　　B. 长期设计和短期设计

C. 全过程设计和单阶段设计　　D. 整体设计和单阶段设计

（4）下列调查属于全面调查的是（　　）。

A. 到某大豆产地了解油料作物的生产情况

B. 就全国宝钢、鞍钢、首钢等重点钢铁生产企业进行调查

C. 对一批出口水果罐头的质量进行抽查

D. 对某市全部工业企业生产经营情况的调查

（5）连续调查与不连续调查的划分依据是（　　）。

A. 调查的组织形式不同　　B. 调查登记的时间是否连续

C. 调查单位包括的范围是否全面　　D. 调查资料的来源不同

（6）对某地区饮食从业人员的身体健康状况进行调查，调查对象是该地区饮食业的（　　）。

A. 全部网点　　B. 每个网点　　C. 所有从业人员　　D. 每个从业人员

3．多选题

（1）专门调查具体包括（　　）。

A. 普查　　B. 统计报表　　C. 重点调查　　D. 抽样调查

E. 典型调查

（2）下列论述中正确的有（　　）。

A. 普查和统计报表都是全面调查

B. 重点调查、典型调查、抽样调查都是非全面调查

C. 经常性调查都是定期调查

D. 一次性调查都是不定期调查

E. 抽样调查是我国收集资料的主要方法

（3）下列情况的统计调查，属于经常性调查的有（　　）。

A. 商品销售额

B. 运输部门货物周转量

C. 每年新出生人口数

D. 某高校毕业生数

E. 突发性自然灾害造成的后果与善后处理

（4）下列情况的统计调查，属于一次性调查的有（　　）。

A. 某地区耕地面积

B. 某县的粮食总产量
C. 学校年底在校学生数
D. 某工业局企业数
E. 企业拥有的固定资产总额

（5）抽样调查方式的优越性表现在以下几个方面（　　）。

A. 全面性　　B. 经济性　　C. 时效性　　D. 准确性
E. 灵活性

（6）统计调查的方式主要有（　　）。

A. 普查　　B. 重点调查　　C. 典型调查　　D. 抽样调查
E. 统计报表

4．计算题

（1）对50只灯泡的耐用时数进行测试，所得数据如下（单位：小时）：

886	928	999	946	950	864	1 050	927	949	852
1 027	928	978	816	1 000	918	1 040	854	1 100	900
866	905	954	890	1 006	926	900	999	886	1120
893	900	800	938	864	919	863	981	916	818
946	926	895	967	921	978	821	924	651	850

要求：

1）试根据上述资料编制次（频）数分布数列。
2）编制向上或向下累计频数、频率数列。
3）根据所编制的次数分布数列绘制直方图、折线图与曲线图。
4）根据所编制的累计频数、频数数列绘制累计曲线图。
5）根据累计曲线图，指出灯泡耐用时数在1 000小时以上的有多少？占多大比重？灯泡耐用时数在900小时以下的有多少？占多大比重？
6）根据频数分布曲线图说明灯泡耐用时数的分布是属于哪一种类型？

（2）试将所在学校的某个现实问题作为选题，进行一次调查，调查单位不少于30个，写出调查方案，将收集的数据进行整理，编制成统计表和统计图，并简要说明其分布特征。

CHAPTER 3

第 3 章　统计数据的描述

章首引例

常州嘉奕运输有限公司成立于 2002 年，是一家提供三类危险品运输和仓储的物流企业。该公司引进国内目前最先进的保温运输车辆以及完善的客户质量保障体系，为 Asland 化学品（中国）公司、华日新材、上海住友商事以及江苏常隆化工有限公司等国内外知名企业提供高质量的物流服务。近年来，由于合理的服务价格以及高质量的客户服务，公司营运规模每年以很快速度增长，现在已经成长为长江中下游地区有名的三类危险品物流企业。

在这家年轻的公司日常经营管理中，现金流管理是最重要一环。能否确保充足的现金流进行资金周转意味着业务成功与失败。而现金管理的重要环节是对公司应收账款的分析与控制。通过权衡和测度应收账款的平均账龄，管理人员可以准确预测可用现金。公司设定下列目标：发票的超期时间平均不能超过 45 天。超过 60 天的发票价值总额不能超过所有应收款项的 5%。

在最近一次应收账款状况汇总中，根据应收账款的账龄提供了下列描述性统计数据：

平均数　40 天
中位数　35 天
众数　　31 天

这些统计数据表明，发票的平均账龄是 40 天。中位数表明有一半的发票小于 35 天。众数 31 天是最常见的发票账龄。统计数据也显示所有应收账款只有 3%超过 60 天。基于这些信息，管理人员感到满意：应收账款和现金流入量都在控制范围中[①]。

在本章中，你可以了解到如何计算和解释常州嘉奕所使用的一些统计测度。除了平均数、众数、中位数以外，你还可以了解其他一些测量数据离散程度，如全距、平均差、方差和标准差等统计指标。这些数值的测度有助于理解和解释数据。

本章学习目标

- ☑ 了解总量指标的概念、作用与种类；
- ☑ 了解相对指标的概念与作用，掌握各种相对指标的计算方法及其运用应注意的问题；
- ☑ 理解平均指标与标志变异指标的概念与作用，掌握各种平均指标与标志变异指标的计算方法和应用场合。

① 作者感谢常州嘉奕运输有限公司吴志强总经理提供的这个统计实例。

3.1 总量指标

3.1.1 总量指标的概念和作用

总量指标是统计资料经过汇总整理后得到的反映总体在一定时间、一定地点的总规模和总水平的总和指标，其表现形式是具有计量单位的绝对数，因此也称为绝对数指标。例如，2008 年，中国国民经济继续保持平稳较快增长，国内生产总值超过 30 万亿元；物价总水平涨幅得到控制；财政收入 6.13 万亿元；粮食总产量 52 850 万吨，创历史最高水平；进出口贸易总额 2.56 万亿美元；实际利用外商直接投资 924 亿美元，等等。以上指标均为总量指标。总量指标数值的大小受总体范围的制约，即总体范围大，指标数值就大；总体范围小，指标数值就小。

总量指标是统计中最基本的指标，在社会各领域的研究和管理中有重要作用。

- 总量指标是对现象总体认识的起点。这些指标是描述一个国家国情和国力、一个地区或一个企业单位的人力、物力、财力的基本数据。例如，国内生产总值、钢铁产量、石油产量、发电量、汽车产量、进出口贸易总额等总量指标，可以表明一个国家或地区的生产水平和经济实力，还可以用做国际间、地区间经济实力的比较分析。只有了解了一个企业的总产值、职工总人数、固定资产总额、厂房总面积、机器设备总台数等总量指标，才能对该企业有基本的认识。
- 总量指标使我们能正确把握情况，是制定政策、编制计划、进行科学管理的重要依据。例如，一个国家的城乡居民储蓄存款余额、全社会固定资产投资总额、货币流通量等总量指标是国家制定货币发行量、存贷款利率、存贷款额度、基本建设投资规模等各项金融政策和财政政策的基础。
- 总量指标是计算相对指标和平均指标的基础。相对指标和平均指标一般是两个有联系的总量指标对比的结果，它们是总量指标的派生指标。如果总量指标计算不科学，统计范畴不合理，将直接影响相对指标和平均指标的准确性。

3.1.2 总量指标的种类

1．总量指标按其反映总体的内容不同，分为总体单位总量和总体标志总量

总体单位总量表明总体在一定时间、地点条件下所包含的个体单位的总数，反映的是总体达到的总规模，如商店数、企业数、学校数、职工人数、学生数等。总体标志总量是指总体内各单位某一标志值的总和，表明总体的某一特征在一定时间、地点条件下达到的总水平，如总产量、总产值、工资总额等。一个总量指标究竟是总体标志总量还是总体单位总量随研究目的的不同和研究对象的变化而定。比如，医生人数这一总量指标，当医院

作为研究总体时，它就是总体标志总量；如果医生作为总体，它就是总体单位总量。

2．总量指标按其反映总体的时间状态不同，分为时期指标和时点指标

时期指标表明总体在一段时间内累积的总量，如产品的产值、国内生产总值、商品销售额等。

时期指标的特点有：

- 不同时期的指标数值具有可加性，相加后表示较长时期现象总的发展水平。
- 时期指标数值大小与所包含的时期长短有直接关系，一般情况下，所包含的时期越长，指标数值越大，时期越短，指标数值越小。比如，一年的总产值总比某一月的产值大。
- 时期指标是连续登记、累计的结果。

时点指标表明总体在某一时刻的数量状态，如人口总数、资金占用余额、库存总额等。

时点指标的特点有：

- 不同时点的指标数值不具有可加性，相加后不具有实际意义。
- 时点指标数值大小与其时间间隔（两个不同时点指标之间的时间距离）长短没有直接关系。比如，年底时某企业的职工人数未必比某一月底职工人数多。
- 时点指标数值是间断计数的。

在实际工作中，这两种总量指标在统计分析和计算处理的方法上都有所不同，须加以区别。

3．总量指标按其计量单位不同，分为实物量指标、价值量指标和劳动量指标

实物量指标可以反映现象总体的使用价值总量。它根据现象的自然属性和特点采用实物单位计量，其实物单位有自然单位、度量衡单位和标准实物单位等。自然单位如人口以“人”为单位，汽车以“辆”为单位，牲畜以“头”为单位等计量。度量衡单位是以统一的度量衡制度规定的单位计量，如钢材以“吨”、木材以“立方米”、建筑面积以“平方米”为单位等计量。此外，实际工作中还经常用标准实物单位计量，标准实物单位用于加总不同规格同类事物的实物数量，可以更加准确地反映产品的使用价值量。例如，不同大小的拖拉机，其发动机的功率大小不一。还有些现象用一种计量单位较难正确反映其使用价值，因此把两种计量单位结合起来，称为复合单位，如货物运输量以“吨公里”为计量单位。不同的实物指标表示不同的质，因此不同的实物指标不能相加。

价值量指标用价值单位反映现象总体的价值总量，具体用货币单位表示。描述生产经营综合成果、社会财富以及进行国民经济核算和企业经济核算的总量指标都采用价值量指标，如“国内生产总值”、“工资总额”、“商品销售总额”等。价值量指标能将任何种类、用途的产品或商品数量加总，因为不同现象之间的价值量指标没有质的差异，所以具有可

加性，应用也最为广泛。

劳动量指标是用劳动时间为单位计量的产品产量或完成的工作量，通常用于工业企业内部的核算，如“工时”、“工日”、“台时”。工时是工人数和工作时数的乘积；工日是工人数和工作日数的乘积；台时是设备台数和开动时数的乘积，也是一种复合单位。企业根据具体条件制定生产单位产品或完成单位作业量所需要的时间标准，即工时定额。按照这种工时（或工日）定额计算产品总量或完成的作业总量，便是劳动量总量指标，它的优点是可以把不同种类、规格的产品产量或作业量进行加总。企业常用这种定额工时产量指标安排作业计划和核定工人的劳动成果。

3.2 相对指标

3.2.1 相对指标的概念与作用

相对指标是用两个有联系的指标进行对比的比值来描述现象数量特征和数量关系的综合指标，常用来反映现象的发展程度、结构、强度、普遍程度或比例关系。相对指标以相对数形式表示，也称做相对数或相对量。

相对指标有以下几个作用：

- 说明总体内在的结构特征为深入分析事物的性质提供了依据。例如，一个地区的第一产业、第二产业、第三产业增加值之间的比例，可以说明该地区经济的发展情况；一个地区城镇人口占该地区总人口的比重，可以说明该地区城市化水平；等等。
- 将现象的绝对差异抽象化，使原来不能直接相比的总量指标可以进行比较。由于不同企业的生产规模、生产条件、职工人数不同，直接用总产值或利润总额等总量指标进行比较与评价意义不大，但用各自的计划完成程度、资金利润率、资金产值率、发展速度等相对指标进行比较，就可以对其生产经营的结果做出较为合理的评价。
- 说明现象的相对水平，表明现象的发展过程和程度，反映事物发展变化的趋势。例如，人均国民收入、人均粮食产量、人均钢铁产量、每千人医生数量、每百人大学生数量等相对指标，可以反映一个国家或地区的国情国力，表明经济实力和医疗、教育的相对水平，发展速度则可以揭示经济发展变化的趋势和方向等。

相对指标的计量形式可以分为两大类：一类是无名数，它是被抽象化的数值，大多采用系数（倍数）、成数、百分数和千分数等表示；另一类是有名数，它把计算相对指标的分子和分母的计量单位同时使用，形成复合单位，如人均国民收入就是用“元 / 人”来表示，人口密度是以“人 / 平方公里”表示等。

3.2.2　相对指标的种类及计算方法

在各实际应用领域中，相对指标按其作用和计算方法不同可分为结构相对指标、比例相对指标、比较相对指标、强度相对指标、动态相对指标与计划完成相对指标 6 种。

1．结构相对指标

结构相对指标是将总体按某一标志分组，然后将各组指标数值与总体指标数值对比的比值，反映总体内部的构成，也叫比重指标。一般用百分数或成数表示。其计算公式为：

$$\text{结构相对指标}=\frac{\text{各组(或部分)总量}}{\text{总体总量}}\times 100\% \qquad (3\text{-}1)$$

⇨【例 3-1】

我国 2007 年国内生产总值为 249 529.9 亿元，其中第一产业增加值 28 095.0 亿元，占国内生产总值的比重为 11.3%，第二产业增加值 121 381.3 亿元，占国内生产总值的比重为 48.6%，第三产业增加值 100 053.5 亿元，占国内生产总值的比重为 40.1%。和世界上发达国家的第一产业占国内生产总值的比重不到 10%，而第三产业占国内生产总值的比重高达 60%相比，差距还是较大的。

由此可见，在研究社会经济现象时，结构相对指标具有重要意义。它经常用来分析现象总体的内部构成情况，说明事物的性质和特征。把不同时间的结构相对指标进行对比分析，可以说明现象的变化过程规律。随着经济建设的不断发展，我国人口城市化程度越来越高，如表 3-1 所示。

表 3-1　我国五次人口普查城乡人口比例分布表（%）

普查次数	普查时点	全国	城市	乡村
第一次人口普查	1953 年 7 月 1 日零时	100	13.26	86.74
第二次人口普查	1964 年 7 月 1 日零时	100	18.40	81.60
第三次人口普查	1982 年 7 月 1 日零时	100	20.60	79.40
第四次人口普查	1990 年 7 月 1 日零时	100	26.20	73.80
第五次人口普查	2000 年 11 月 1 日零时	100	36.09	63.91

2．比例相对指标

比例相对指标是描述总体内部各个组成部分之间的数量对比关系的相对指标。一般用百分数表示，也可用一比几或几比几形式表示，分析总体中若干部分的比例关系时可采用连比形式。其计算公式为：

$$比例相对指标=\frac{总体中某一部分数值}{总体中另一部分数值}\times 100\% \qquad (3\text{-}2)$$

在表 3-1 中，我国第五次人口普查的结果显示，城市人口与乡村人口的比例可以表示为 36.1∶63.9。再如，例 3-1 中的第一、第二和第三产业的增加值的比为 1∶3.5∶2.1。

比例相对指标能够反映总体内部各部分之间的数量联系程度。社会经济中的许多重大比例关系，如人口的性别比例关系、积累和消费的比例关系、产业的比例关系，都可以用比例相对指标来反映。分析研究这些比例关系，有助于发现社会经济现象中的规律，调整不合理的比例，促使社会经济协调发展。

3．比较相对指标

比较相对指标是将某一总体的指标与另一总体同类指标对比的相对指标。一般用倍数或百分数表示。其计算公式为：

$$比较相对指标=\frac{甲单位某指标值}{乙单位同类指标值}\times 100\% \qquad (3\text{-}3)$$

➪【例 3-2】

甲乙两个上市公司 2008 年的利润总额分别为 5 亿元和 2 亿元，则甲公司的利润是乙公司的 2 倍。

计算比较相对指标是用总量指标来对比的，常常由于总体的规模和条件的影响，使结果不能正确反映现象发展的本质差异，所以可采用相对指标和平均指标进行比较，即计算每股利润指标。通过计算，甲公司的每股赢利为 0.2 元，而乙公司的每股赢利是 0.5 元，那么乙公司的每股赢利是甲公司的 2.5 倍。

由此可见，公司的规模是甲公司大，但赢利能力却是乙公司强。

比较相对指标是反映不同国家、不同地区或不同单位之间同类现象的差异程度，所以要注意对比的两个同类指标数值必须具有可比性，即在指标含义、口径、计算方法、计量单位、所属时间等方面要一致。比较相对指标比较的基数不固定，可以视不同的研究目的而定，一般而言，应以主要观察单位的指标数值作为对比基数，以反映与其他单位的差距。

4．动态相对指标

动态相对指标是将总体不同时期的同一类指标对比而计算的比值，说明现象水平发展变化的方向和程度。一般用百分数表示。通常，将作为比较基础的时期称为基期，与基期对比的时期称为报告期。其计算公式为：

$$\text{动态相对指标}=\frac{\text{报告期指标值}}{\text{基期指标值}}\times 100\% \qquad (3\text{-}4)$$

动态相对指标的具体内容将在第 6 章中详细阐述。

5．强度相对指标

强度相对指标是指两个性质不同但有一定联系的总量指标之比。它用来表明现象的强度、密度和普遍程度等，以表示不同现象之间依存性比例关系的综合指标。其计算公式为：

$$\text{强度相对指标}=\frac{\text{某一总量指标数值}}{\text{另一性质不同而有联系的总量指标数值}}\times 100\% \qquad (3\text{-}5)$$

强度相对指标一般采取复合单位表示，如平均每人粮食产量为“公斤／人”，平均每人钢铁产量为“吨／人”，每百人中大学生的人数为“人／百人”等；也有采用非复合单位的，如商品流转次数用“次”表示；还有既可用复合单位表示，也可用百分数、千分数表示的，如资金利税率、商品流通费用率等可用“元／百元”或百分数表示，而人口死亡率用千分数表示。

⇨【例 3-3】

我国 2007 年国内生产总值为 249 529.9 亿元，全年粮食产量为 50 160.3 万吨，城乡居民储蓄存款为 389 371 亿元，2004 年年末全国人口为 132 129 万人，则：

我国人口密度 = 132 129 万人 /960 万平方公里 =137.6（人 /平方公里）

我国人均国内生产总值 =249 529.9 亿元 /132 129 万人 =18 885.3（元 /人）

我国人均粮食产量 =50 160.3 万吨 /132 129 万人 =0.38（吨 /人）

我国人均银行存款余额 =389 371 亿元 /132 129 万人 =29 469（元 /人）

强度相对指标还有正指标、逆指标之分。正指标比值的大小与其反映的强度、密度和普遍程度成正比，逆指标比值的大小与其反映的强度、密度和普遍程度成反比。如上例中的各项指标均为正指标。有些强度相对指标将其比式的分子分母互换，就可从正指标变为逆指标，或者由逆指标变为正指标，其评价判别的意义相同。

如上例中的人口密度还可用逆指标表示，即：

我国人口密度 = 960 万平方公里 /132 129 万人= 0.007 27（平方公里 /人）

6．计划完成相对指标

计划完成相对指标是将某一时期的实际完成数与同期计划数进行对比，反映计划执行情况的相对指标，一般用百分数表示。计算公式如下：

$$计划完成相对指标=\frac{某期实际完成指标值}{某期计划完成指标值}\times 100\% \qquad (3\text{-}6)$$

实际工作中，由于计划数可表现为绝对数（总量指标）、相对数（相对指标）、平均数（平均指标）等多种形式，因此计算计划完成相对指标的方法也不尽相同。

（1）计划数为绝对数（总量指标）、平均数（平均指标）

计划完成相对指标的计算公式为：

$$计划完成相对指标=\frac{实际完成总量（或平均数）}{同期计划总量（或平均数）}\times 100\% \qquad (3\text{-}7)$$

它一般适合考核现象的规模或水平的计划完成情况。

在检查中长期计划（如五年计划）任务的完成情况时可将指标分为水平法和累计法两种。

- 水平法是用来检查中长期计划的末期是否达到了规定的水平，适用于反映生产能力的经济指标，如钢产量、水泥产量、石油产量、煤产量、发电量等指标的计划完成情况的检查。它以计划期期末水平（即最后一年水平）为考核对象，计算公式为：

$$计划完成相对指标=\frac{计划期末实际达到水平}{计划期末规定的期末水平}\times 100\% \qquad (3\text{-}8)$$

计算提前完成计划的时间的方法是：在计划期内，只要有连续12个月的实际完成数（可以跨年度）达到计划规定的末年水平就算完成了计划。剩余的时间就是提前完成计划的时间。

⇨【例3-4】

某地区2004年至2008年规定某种产品达到年产80万吨的水平，实际在计划末期，即2008年已达到86万吨，则该产品产量计划完成程度为86/80×100%=107.5%。表明超额7.5%完成了计划。

若该产品产量从2007年的8月起至2008年7月实际已达到80万吨，则该产品产量提前5个月完成了五年计划任务。

- 累计法用来检查计划期内的若干年累计完成量是否已达到应有的水平，一般适用于构成国民财产存量的经济指标，如固定资产投资、住宅建设、开耕荒地等计划完成情况的检查。它以计划期内各年计划数量的累计总和为考核对象，计算公式为：

$$计划完成相对指标=\frac{计划期间实际完成累计数}{计划期规定的累计数}\times 100\% \qquad (3\text{-}9)$$

计算提前完成计划的时间的方法是：从计划期开始至某一时间，累计的实际完成数已达到计划数，就算完成了计划。将计划全部时间减去完成计划所需的时间就为提前完成计划的时间。



➪【例 3-5】

某地区“十五”期间规定五年累计造林面积 3 000 亩，而实际完成造林面积累计达 3150 亩，则该地区造林面积“十五”计划完成程度为 3150／3 000×100%=105%。表明超额 5%完成了计划。

若截至 2008 年 10 月底已累计造林 3 000 亩，则五年计划提前两个月完成。

（2）计划数为相对数（相对指标）

计划完成相对指标的计算公式为：

$$\begin{aligned}\text{计划完成相对指标}&=\frac{\text{实际完成数（\%）}}{\text{同期计划数（\%）}}\times 100\%\\&=\frac{100\%+\text{实际提高率（或}-\text{实际降低率）}}{100\%+\text{计划提高率（或}-\text{计划降低率）}}\times 100\%\end{aligned}\qquad(3\text{-}10)$$

➪【例 3-6】

某工厂的某种产品的产量计划要求增长 8%，单位成本计划要求下降 4%，而实际产量增长了 12%，实际单位成本下降了 5%，则计划完成程度指标为：

$$\text{产量计划完成相对指标}=\frac{100\%+12\%}{100\%+8\%}\times 100\%=103.7\%$$

$$\text{单位成本计划相对指标}=\frac{100\%-5\%}{100\%-4\%}\times 100\%=98.96\%$$

计算结果表明：该企业产量计划完成了 103.7 %，超额完成了 3.7 %；单位成本计划完成了 98.96%，超计划完成 1.04%。这说明产量比计划提高近 4 个百分点，单位产品成本比计划降低 1 个百分点。

3.3 集中趋势的描述 —— 平均指标

3.3.1 平均指标的概念

在社会经济现象的同质总体中，同一标志在各单位的数量表现不尽相同，标志值大小各异，这就需要利用平均指标来代表总体的一般水平。平均指标，是同类社会经济现象总体内各单位某一数量标志在一定时间、地点和条件下的一般水平，是总体内各单位参差不齐的标志值的代表值。其数值表现为平均数。平均指标是社会经济统计中常用的综合指标之一。

平均指标固然决定于总体内各单位个体的水平，但它反映的是总体的数量特征，是总体变量分布的一个重要特征值。无论是自然现象或社会经济现象，很多变量的分布都表现为接近平均数的标志值居多，远离平均数的标志值较少，也即多数标志值以平均数为中心密集地分布在它的两侧，呈现出向心力作用下的集中趋势。因此，平均指标也是对变量分布集中趋势的测定，反映分布集中趋势的特征。

3.3.2 平均指标的计算方法与应用

社会经济统计中的平均指标有算术平均数、调和平均数、几何平均数、中位数和众数五种。前三种平均数是根据总体全部单位标志值计算的，又称之为数值平均数；后两种平均数是根据其所处位置有关的部分标志值计算的，又称之为位置平均数，在某些特定场合，可代替数值平均数来反映现象的一般水平。

1. 算术平均数

算术平均数是平均指标中应用最基本、最广泛的一种。其基本计算公式是：

$$\text{算术平均数}=\frac{\text{总体标志总量}}{\text{总体单位总量}} \tag{3-11}$$

在社会经济现象中，总体的标志总量常常是总体单位标志值的算术总和。

例如，工人工资总额是各个工人工资的总和；粮食总产量是各块地播种面积产量的总和；等等。因此，这些现象的平均值，一般都是该现象的标志总量与单位数之比。

利用该基本公式计算平均数时，要注意公式的子项与母项必须属于同一总体，分子必须是分母中各单位标志值之和，分母必须是总体标志总量中标志值的个数。否则，计算的平均数就会失去实际意义。这也是平均指标与强度相对指标的区别之一。

由于掌握的资料不同，算术平均数可分为简单算术平均数和加权算术平均数两种。

（1）简单算术平均数

简单算术平均数就是直接将总体各单位的标志值相加，除以总体单位数而求得平均数，一般应用于资料未经过任何分组整理与加工的情况。其计算公式为：

$$\bar{x}=\frac{x_1+x_2+\mathrm{L}+x_n}{n}=\frac{\sum_{i=1}^{n}x_i}{n}\text{或}\frac{\sum x}{n} \tag{3-12}$$

式中，$\bar{x}$ 是简单算术平均数；x_i 是各单位标志值；n 是总体单位数；$\sum$ 是总和符号。

【例 3-7】

某厂某生产小组有 10 名工人，月工资分别为：800，800，700，620，700，1 060，800，700，800，620 元，则该 10 名工人平均月工资为：

$$\bar{x}=\frac{800+800+700+620+700+1\,060+800+700+800+620}{10}=760(元)$$

简单算术平均数的大小只受变量值大小的影响，不受总体单位数多少的影响。

（2）加权算术平均数

在资料已经分组得出次数分配的条件下计算平均数，要先将各组的标志值与其相应的次数相乘，求出各组的标志总量，再把各组的标志总量相加，求出总体的标志总量，再除以总体单位总数。按这种方法计算的平均数就叫做加权算术平均数。其计算公式为：

$$\bar{x}=\frac{x_1f_1+x_2f_2+\mathrm{L}\ +x_nf_n}{f_1+f_2+\mathrm{L}\ +f_n}=\frac{\sum_{i=1}^{n}x_if_i}{\sum_{i=1}^{n}f_i}\ 或\ \frac{\sum xf}{\sum f} \qquad (3\text{-}13)$$

式中，$\bar{x}$ 是加权算术平均数；f 是各组单位数，即次数；x_i 是各组标志值；$\sum$ 是总和符号。

分组资料有单项式变量数列和组距式变量数列，下面分别举例说明。

1）由单项数列计算加权算术平均数

【例 3-8】

仍用上例资料，将该资料进行分组整理成单项数列，如表 3-2 所示。

表 3-2　某厂生产小组工人月工资

按月工资额分组 x（元）	工人人数 f（人）	比重 $f/\sum f$
620	2	0.20
700	3	0.30
800	4	0.40
1 060	1	0.10
合 计	10	1.00

该生产小组工人平均月工资为：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{620\times2+700\times3+800\times4+1\,060\times1}{3+2+4+1}=\frac{7\,600}{10}=760（元）$$

按加权算术平均数的方法计算平均数，平均数的大小不仅取决于总体各单位标志值（x）的大小，而且还受到各标志值出现次数（f）多少的影响。人数多的组，其标志值对平均数的影响大；人数少的组，其标志值对平均数的影响小。即当标志值比较大的组的次数多时，平均数就接近大的一方；当标志值比较小的组次数多时，平均数就接近标志值小的一方。标志值的次数（f）多少对平均数（$\bar{x}$）大小的影响具有权衡轻重的作用。因此，在统计中通常把各组单位数称为权数，把每个标志值乘以权数的过程叫加权过程，这样计算的算术平均数又叫加权算术平均数。

必须指出，权数对于算术平均数的影响作用，就其实质而言，并不是决定于各组单位数次数的多少，而决定于各组单位数占总体单位数的比重（又称为权重系数）的大小。即，哪一组单位数所占比重大，哪一组标志值对平均数的影响就大。对式（3-13）作变换，就可以清楚地看到。

$$\bar{x}=\frac{x_1f_1+x_2f_2+\cdots+x_nf_n}{f_1+f_2+\cdots+f_n}=\frac{x_1f_1}{\sum f}+\frac{x_2f_2}{\sum f}+\cdots+\frac{x_nf_n}{\sum f}=\sum\left(x\frac{f}{\sum f}\right)$$

如上例中，也可这样计算

$$\bar{x}=\sum\left(x\frac{f}{\sum f}\right)=620\times0.2+700\times0.3+800\times0.4+1\ 060\times0.1=760\text{（元）}$$

从以上计算可以看出，用总体单位数权数和总体单位数比重权数的计算结果是一致的。由此可见，总体单位数的比重是权数起作用的实质，如果仅仅是总体单位数发生了变化，而各组单位数的比重未变，则平均数不会改变。

2）由组距数列计算加权算术平均数

其方法基本上相同，只要先计算出各组的组中值，再以各组的组中值代表该组的标志值，然后再来计算加权算术平均数。

【例3-9】

某企业某月工人工资情况如表3-3所示。

表3-3　某企业某月工人工资加权算术平均数计算表

按月工资数分组（元）	组中值x（元）	职工工人数f（人）	各组职工工资总额xf（元）
600以下	500	10	5000
600～800	700	20	14000
800～1 000	900	35	31500
1000～1200	1100	25	27500
1 200以上	1300	10	13000
合　计	—	100	91000

职工月平均工资为：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{91\ 000}{100}=910\text{（元）}$$

应当说明，根据组距数列计算算术平均数的方法具有一定的假定性，即假设各单位标志值在各组内的分布是均匀的，而组中值是各组标志值的代表值。实际上，各组内部的标志值变动不一定是完全均匀的。因此，由组中值计算的加权平均数的数值是实际平均数的近似值，组距越小，组中值的代表性越高，计算得到的平均数就越接近实际的平均数。

（3）算术平均数的几个主要数学性质

为了在实际应用中更好地计算与分析平均数，我们有必要了解算术平均数的性质。

- 平均数与次数的乘积等于变量值与次数的乘积的总和，即：

$$n\bar{x}=\sum x \quad 或 \quad \bar{x}\sum f=\sum xf$$

这个性质说明，平均数是所有变量值的代表数值，并且根据平均数次数可以推算出数量标志的总和。

- 各个变量值与算术平均数的离差之和等于零，即：

$$\sum(x-\bar{x})=0 \ 或 \ \sum(x-\bar{x})f=0$$

在理论上，这个性质说明，在算术平均数中，变量之间的偏差可以相互抵消。

- 各个变量值与平均数离差平方之和为最小，即：

$$\sum(x-\bar{x})^2=\text{最小值} \ 或 \ \sum(x-\bar{x})^2 f=\text{最小值}$$

这个性质有助于理解方差和均方差的意义。

2．调和平均数

调和平均数是同质总体中，各单位标志值倒数的算术平均数的倒数，故又称为倒数平均数。

设某同质总体有 n 个变量值 x_1，x_2，…，x_n，由定义得各个变量值的倒数为：

$$\frac{1}{x_1},\frac{1}{x_2},\cdots,\frac{1}{x_n}$$

变量值倒数的算术平均数为：

$$\frac{\dfrac{1}{x_1}+\dfrac{1}{x_2}+\cdots+\dfrac{1}{x_n}}{n}$$

变量值倒数的算术平均数的倒数为：

$$H=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$$

由于掌握资料的不同，调和平均数也可分为简单调和平均数和加权调和平均数两种。

（1）简单调和平均数

根据对调和平均数概念的理解，可以得出简单调和平均数计算公式如下：

$$H=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\mathrm{L}+\frac{1}{x_i}+\mathrm{L}+\frac{1}{x_n}} \qquad (3\text{-}14)$$

式中，H 是调和平均数；x_i 是各单位标志值（变量值）；n 是标志值项数。

简单调和平均数用于资料未分组的情况下。

➪【例 3-10】

某市场上某种蔬菜早市为每千克 0.50 元，中市为每千克 0.40 元，晚市为每千克 0.20 元。若早、中、晚市各买 1 千克，平均每千克价格为：

$$\bar{x}=\frac{\sum x}{n}=\frac{0.50+0.40+0.20}{3}=0.37\text{（元）}$$

若早、中、晚市各买 1 元钱的菜，则平均每千克价格又是多少呢？在此条件下，作为计算平均指标基础的总体单位数不是 3 千克，而是：

$$\frac{1}{0.5}+\frac{1}{0.4}+\frac{1}{0.2}=9.5\text{（千克）}$$

计算这 9.5 千克蔬菜的平均价格，则应采用简单调和平均数公式：

$$H=\frac{n}{\sum\frac{1}{x}}=\frac{3}{9.5}=0.32\text{（元）}$$

在各买 1 千克的场合下，由于千克数相同，每种价格对平均价格的影响是相等的；在各买 1 元钱的场合下，虽然金额相同，但每种价格不同，所以千克数就不同，也就意味着每种价格对平均价格的影响是不同的。价格最低的晚市蔬菜买的数量最多，以致平均价格相对降低了。调和平均数就是在这种意义上的应用。

（2）加权调和平均数

对于分组资料计算调和平均数，要用加权式。其计算公式为：

$$H=\frac{m_1+m_2+\cdots+m_n}{\frac{m_1}{x_1}+\frac{m_2}{x_2}+\cdots+\frac{m_n}{x_n}}=\frac{\sum m}{\sum\frac{m}{x}} \tag{3-15}$$

式中，m 是权数；$m=xf$ 是各组标志总量。

➪【例 3-11】

某工厂本月购进某种材料四批，每批价格以及采购金额如表 3-4 所示。

表 3-4　某工厂某材料采购情况表

	价格 x（元/千克）	采购金额 m（元）	采购量 $\frac{m}{x}$（千克）
第一批	35	10 000	286
第二批	40	20 000	500
第三批	45	15 000	333
第四批	50	5 000	100
合　计	—	50 000	1 219

已知资料　　计 算 栏

平均每千克价格为：

$$H=\frac{\sum m}{\sum\frac{m}{x}}=\frac{50\,000}{1\,219}=41.02\ （元）$$

该例中，m 为采购金额，即权数；价格 x 为变量值；分子是采购金额，即总体标志总量；分母为采购量之和，即总体单位总数。所以，调和平均数仍然是以总体标志总量除以总体单位总数计算的。它在经济内容和计算结果上与算术平均数一致，只是由于掌握资料的不同，而在计算公式和计算过程方面有别于算术平均数。将 $m=xf$，$f=\frac{m}{x}$ 代入加权算术平均数公式，得：

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{\sum m}{\sum\frac{m}{x}}=H$$

可见，加权调和平均数实际上是加权算术平均数的变形。在实际应用中，若掌握的是变量值和总体单位数的资料，则采用算术平均数公式计算平均数；若掌握的是变量值

和总体标志总量而缺少总体单位资料，就应用调和平均数公式计算平均数。

3．几何平均数

几何平均数就是 n 个变量值连乘积的 n 次方根。

由于掌握资料的差异，几何平均数也分为简单几何平均数和加权几何平均数两种。

（1）简单几何平均数

设有 n 个变量值 x_1，x_2，…，x_n，由几何平均数定义可得出简单几何平均数的计算公式为：

$$G=\sqrt[n]{x_1\cdot x_2\mathrm{L}\ x_n}=\sqrt[n]{\prod x} \qquad (3\text{-}16)$$

式中，G 是几何平均数；$\prod$ 是连乘符号。

➪【例 3-12】

某机械厂生产机器，设有毛坯、粗加工、精加工、装配四个连续作业的车间，某批产品的各车间的合格率分别为 96%，93%，95%，97%，求各车间制品平均合格率。

由于全厂产品的总合格率并不等于各车间制品的合格率总和，后续车间的合格率是在前一车间制品全部合格的基础上计算的。全厂产品的总合格率应等于各车间制品合格率的连乘积，所以不能采用算术平均数和调和平均数公式计算平均合格率，而应用几何平均法来求得。即车间制品平均合格率为：

$$G=\sqrt[n]{\prod x}=\sqrt[4]{96\%\times 93\%\times 95\%\times 97\%}=95.24\%$$

（2）加权几何平均数

当计算几何平均数的每个变量值的次数不相同时，则应用加权几何平均法，其计算公式为：

$$G=\sqrt[f_1+f_2+\mathrm{L}+f_n]{x_1^{f_1}\cdot x_2^{f_2}\mathrm{L}\ x_n^{f_n}}=\sqrt[\sum f]{\prod x^f} \qquad (3\text{-}17)$$

式中，f 是变量值的次数；$\sum f$ 是次数总和；其他符号同前。

➪【例 3-13】

某笔为期 20 年的投资按复利计算收益，前 10 年的年利率为 10%，中间 5 年的年利率为 8%，最后 5 年的年利率为 6%，则 20 年后的本利率为：

$$(1+10\%)^{10}\times(1+8\%)^5\times(1+6\%)^5=5.100\ 1$$

整个投资期间的年平均利率为：

$$G=\sqrt[10+5+5]{1.10^{10}\times 1.08^5+1.06^5}-1=\sqrt[20]{5.100\,1}-1=8.487\%$$

几何平均数是计算平均比率或平均速度最适用的一种方法，这是因为几何平均数的数学性质与社会经济现象发展的平均比率或平均速度形成的客观过程相一致。当一个数列的后一个数据以前一个数据为基础成比率增长、总比率等于这些变量值的连乘积时，都适用于几何平均法计算平均比率。

4．众数

众数（M_o）是在总体中出现次数最多的标志值，是总体中最常遇到的最普遍、最一般的变量值，因而常用来说明社会经济现象的一般水平。例如，为了掌握农贸市场某种商品的价格水平，往往利用该种商品最普遍的成交价格为代表；又如为了满足广大消费者的需要，有关生产与销售部门就必须了解消费者需求量最大的服装、鞋帽的尺码、规格、型号等，作为制定生产和销售计划的重要依据。可见，众数的运用，有其特殊的意义。

一般地说，众数的确定比较简单，只要通过观察就可得知，即以次数最多的那个标志值为众数。但如果掌握的资料是组距数列，众数的确定方法略为复杂一些。

（1）由单项数列确定众数

【例 3-14】

某班同学年龄分组资料如表 3-5 所示。

表 3-5　某班同学年龄分布表

按年龄分组（岁）	人数（人）
18	3
19	8
20	14
21	4
22	1
合　　计	30

经观察发现，20 岁的同学人数最多，则众数为 20（岁）。

（2）由组距数列确定众数

对于组距数列，确定众数需分两步进行：

首先，确定众数组，即从变量数列中找出次数或频率最大的组，该组的上、下限就规定了众数的可能取值范围。

其次，依据与众数组相邻的两个组的次数，利用公式近似计算众数值。

【例 3-15】

某城市居民家庭收入的抽样调查资料如表 3-6 所示。

表 3-6　某城市居民家庭收入情况表

年收入水平（元）	居民户数 f	向上累计居民户数 s	向下累计居民户数 s
10 000 以下	92	92	1 000
10 000～15 000	180	272	908
15 000～20 000	240	512	728
20 000～25 000	260	772	488
25 000～30 000	140	912	228
30 000～35 000	53	965	88
35 000 以上	35	1 000	35
合　　计	1 000	—	—

表中将全部被调查居民户按收入高低分为 7 个组，并列出了各组的户数和累计户数资料。很容易看出，表中的第 4 组（年收入为 20 000～25 000 元的组）是该变量数列的众数组。

具体确定众数值，要根据众数组相邻两组次数的多少而定。当众数组相邻两组次数相等或当该数列次数分布对称时，众数组的组中值就是众数；当众数组前一组的次数大于众数组后一组的次数时，则众数靠近众数组的下限；当众数组前一组的次数小于众数组后一组的次数时，则众数偏向众数组的上限。因此，在众数组内确定众数时，可按众数组次数与其相邻两组次数的差数的比例，采用下限公式或上限公式来计算众数的近似值，这一推算方法称为比例插值法，如图 3-1 所示。

下限公式：

$$M_o = L_{M_o} + \frac{f_{M_o} - f_{M_{o-1}}}{(f_{M_o} - f_{M_{o-1}}) + (f_{M_o} - f_{M_{o+1}})} \times i \qquad (3\text{-}18)$$

上限公式：

$$M_o = U_{M_o} + \frac{f_{M_o} - f_{M_{o+1}}}{(f_{M_o} - f_{M_{o-1}}) + (f_{M_o} - f_{M_{o+1}})} \times i \qquad (3\text{-}19)$$

式中，M_o 是众数；L_{M_o} 是众数组下限；U_{M_o} 是众数组上限；$f_{M_{o-1}}$ 是众数组前一组次数；

f_{M_o} 是众数组次数；$f_{M_{o+1}}$ 是众数组后一组次数；i 是众数组组距。

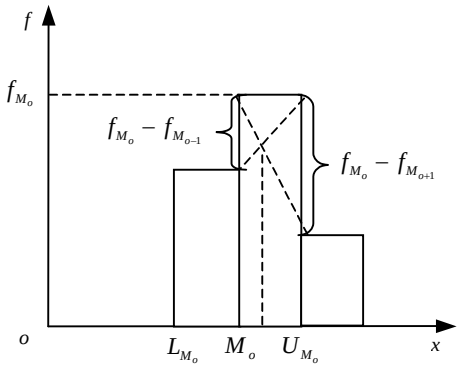


图 3-1　众数计算公式示意图

将表 3-6 中的数据分别代入式（3-18）、式（3-19）得：

$$M_o = 20\,000 + \frac{260-240}{(260-240)+(260-140)} \times 5\,000 = 20\,714.29(\text{元})$$

或

$$M_o = 25\,000 - \frac{260-140}{(260-240)+(260-140)} \times 5\,000 = 20\,714.29(\text{元})$$

众数是一种位置平均数。它不受各项变量值的影响，因此用它作为平均水平的代表值有不足之处。但当数列中存在异常变量值时，它不受极端变量值的影响，要比算术平均数更能确切地代表现象的一般水平。由于众数属于次数最多而又高度集中的数值，因此，它只适用于分布的次数较多且具有明显集中趋势的总体。当变量数列呈均匀分布时，则无众数可言。

5．中位数

把总体单位的某一标志的各个数值，按大小顺序排列，居中间位置的标志值就是中位数（M_e）。这说明，中位数就是将全部总体单位按标志值的大小平分为两个部分，使得总体中有半数单位标志值小于中位数，而另外半数单位标志值则大于中位数。用这样一个中等水平的标志值来反映现象的一般水平，具有非常直观的代表性意义。如要了解某地区职工收入的一般水平，由于该地区职工收入差距悬殊，因此用职工收入的中位数要比用平均收入更能代表职工的实际平均水平。

为了确定中位数，必须将总体各单位的标志资料按大小顺序排列，最好是编制出变量数列，这里有三种情况。

（1）对于未分组的原始资料，首先必须将标志值按大小排序

设排序的结果为：

$$x_1 \leqslant x_2 \leqslant x_3 \leqslant \cdots \leqslant x_n$$

则中位数就可以按下面的方式确定：

$$M_e = \begin{cases} x_{\frac{n+1}{2}} & \text{当}n\text{为奇数} \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{当}n\text{为偶数} \end{cases} \qquad (3\text{-}20)$$

⇨【例 3-16】

如某小组 5 名工人的产量（件）按顺序排列分别为 20，22，24，26，27，中位数位置为（5+1）/ 2=3，则第 3 个位置的日产量 24 件就是中位数。若例中该小组有 6 名工人，日产量（件）依次为 20，22，24，26，27，28，中位数位置为（6+1）/ 2=3.5，则第 3 个和第 4 个位置上的两个标志值算术平均数就是中位数，即（24+26）/ 2=25 件。

（2）由单项数列确定中位数

先计算各组的累计次数（或累计频率），然后按以下方式确定中位数：

$$M_e = \begin{cases} x_{\frac{\sum f+1}{2}} & \text{当}\sum f\text{为奇数} \\ \dfrac{x_{\frac{\sum f}{2}} + x_{\frac{\sum f+1}{2}}}{2} & \text{当}\sum f\text{为偶数} \end{cases} \qquad (3\text{-}21)$$

⇨【例 3-17】

某企业两组工人生产零件数资料如表 3-7 所示，求中位数。

表 3-7　工人按生产零件数分组表

按生产零件数分组（件/日）	工人人数（人）		人数累计（人）	
	甲　组	乙　组	甲　组	乙　组
20	10	15	10	15
21	20	20	30	35
22	30	30	60	65
23	20	32	80	97
24	15	19	95	116
25	5	14	100	130
合　计	100	130	—	—

甲组工人生产零件的中位数位置为$(\sum f+1)/2=50.5$，说明它位于第 50 个与 51 个工人之间，根据累计次数可知，第 50 个与第 51 个工人都在第 3 组，其标志值相同，都为 22 件，因此不需再平均计算就可确定为中位数为 22 件。

乙组的中位数位置为（130+1）/2 = 65.5，说明它位于第 60 与第 61 个工人之间，对照累计次数可知，这两个工人分别属第 3 组与第 4 组，因此应取这两组标志值的算术平均数为中位数，即中位数为（22+23）/2 = 22.5 件。

（3）由组距数列确定中位数

第一步：根据组距数列中的累计次数$\sum f/2$确定中位数所在组，这个组的上下限就规定了中位数的可能取值范围。

第二步：假定中位数组内的各单位是均匀分布的，用插补法按比例推算出中位数的近似值，公式如下：

上限公式
$$M_e = L_{M_e} + \frac{\frac{\sum f}{2} - S_{M_{e-1}}}{f_{M_e}} \times i_{M_e} \tag{3-22}$$

下限公式
$$M_e = U_{M_e} - \frac{\frac{\sum f}{2} - S_{M_{e+1}}}{f_{M_e}} \times i_{M_e} \tag{3-23}$$

式中，M_e是中位数；L_{M_e}是中位数组下限；U_{M_e}是中位数组上限；f_{M_e}是中位数组次数；$S_{M_{e-1}}$是向上累计至中位数组前一组止的次数；$S_{M_{e+1}}$是向下累计至中位数组后一组止的次数；其他符号同前。

➪【例 3-18】

根据表 3-6 所给出的居民家庭的收入资料可知，该组距数列的中位数组为第 3 组（年收入为 15 000～20 000 元的组）；进一步计算居民家庭年收入的中位数近似值，依据下限公式有：

$$M_e = 15\,000 + \frac{500-272}{240} \times 5\,000 = 19\,750\text{（元）}$$

依据上限公式则有：

$$M_e = 20\,000 - \frac{500-488}{240} \times 5\,000 = 19\,750\text{（元）}$$

中位数也是一种位置平均数，它是数列中间一项或二项标志值的平均数，不受极端值的影响。因此，当数列中存在异常值（极大值或极小值）时，采用中位数能更好地反映现象的一般水平。社会经济统计中，对一些不能用数量表示，而只能用等级、名次等表示的现象，可采用中位数来代表其一般水平。

6．算术平均数与众数、中位数的关系

算术平均数与众数、中位数之间有着一定的关系，这种关系决定于总体内次数分布的状况。当总体呈对称的钟形分布时，算术平均数位于次数分布曲线的对称点上，而该点又是曲线的最高点和中心点。因此，算术平均数、众数和中位数三者相等。当总体呈非对称分布的钟形分布时，算术平均数、众数和中位数之间存在一定的差别。当次数分布右偏时，有 $\bar{x} > M_e > M_o$；当次数分布左偏时，有 $\bar{x} < M_e < M_o$，其差别大小与偏斜程度大小成正比。产生这种差别的原因，是由于这三种平均数受极端变量值影响的程度不同。算术平均数受全部变量值的影响，而且受极端值的影响最大。中位数只受极端值的位置影响而不受其数值影响。众数不受极端值位置也不受其数值影响。因此，当分布呈非对称钟形状态时，算术平均数或小于众数，或大于众数，而中位数总是位于两者之间。三者关系如图 2-3 所示。

3.3.3 正确运用平均指标的原则

为了保证平均指标能正确地表明现象的本质及其发展规律，运用平均指标必须遵守如下三个原则。

1．平均指标必须应用于同质总体

平均数的特点是用一个有代表性的数值反映现象的一般水平。因此，所有个体单位必须在一个同质总体之内，这样计算的平均数，才有代表意义。

2．使用平均指标应和次数分布相结合

在使用平均指标了解现象时，不能只看平均数，要同时看变量数列的次数分布情况和最高标志值与最低标志值。如按表 3-2 的资料计算的平均月工资为 760 元，这是一个代表值，从例 3-9 中还了解到 10 个工人中工资最高为 1 060 元，最低为 620 元，半数工人月工资在 800 元以下，这样才能深刻地了解全面的情况。

3．平均指标与分组法相结合

在动态对比中，利用分组法分析总体结构的变动对平均指标变动的影响；在静态对比中，利用分组法分析双方结构不同对平均指标的影响，从而加深对平均指标的认识。



➪【例 3-19】

企业某车间有三个生产同种产品的班组，其有关资料如表 3-8 所示。

表 3-8　各班组生产同种产品的原材料消耗情况

班组	第 3 季度				第 4 季度			
	产量（件）	百分比（%）	原材料消耗量（千克）	单耗（千克/件）	产量（件）	百分比（%）	原材料消耗量（千克）	单耗（千克/件）
甲	300	54.5	15 000	50	100	16.7	5 300	53
乙	150	27.3	6 750	45	300	50.0	14 100	47
丙	100	18.2	4 400	44	200	33.3	9 000	45
合 计	550	100.0	26 150	47.5	600	100.0	28 400	47.3

从表 3-8 中总单耗来看，第 4 季度为 47.3 千克/件，第 3 季度为 47.5 千克/件，第 4 季度低于第 3 季度。这样能否断定第 4 季度原材料的消耗情况好于第 3 季度呢？这就需要从分组单耗来分析了。从各班组单耗来看，无论甲、乙、丙的单耗都是第 3 季度低于第 4 季度。那么为什么会出现如此矛盾的情况呢？原因在于单耗较低的乙、丙两组产量占总产量的百分比由第 3 季度的 45.5%上升到第 4 季度的 83.3%，而单耗较高的甲组产量百分比由 54.5%下降到 16.7%。因此，第 4 季度和第 3 季度相比，总单耗下降的结果是由于提高了单耗低的组产量占总产量的百分比，而降低了单耗高的组产量占总产量的百分比的结果。

CASE 【实例】　正确理解与运用平均数

无为县农村贫困调查与思考

有关资料显示：2002 年，无为县对国内生产总值的贡献、固定资产投资、城乡居民储蓄存款余额三项经济指标高居全省 61 个县首位。在外人眼里，无为县是经济大县，是经济强县。这样的县还是不是贫困县？否认者有之，怀疑者有之，无为人理不直气不壮。“没有调查就没有发言权”。最近，县扶贫办与农调队对此开展一次入户调查，试图通过事实答疑解惑，以期为领导决策提供服务。带着这一问题，县扶贫办与县农调队从 6 月开始，花了一个月时间，深入到全县 30 个乡镇中 30 个村 300 户，对 2002 年农村贫困问题进行了一次抽样调查。调查结果表明：我县依然相当贫困，扶贫开发任重道远。

贫困特征分析如下：

- 数量多。样本户中，2002 年人均纯收入低于 625 元的绝对贫困户有 68 人、17 户，占调查人口的 5.7%；625～865 元的低收入户有 106 人、28 户，占调查人口的 8.4%；贫困人口占调查人口的 1/7 强。据此测算：2002 年，全县绝对贫困人口 73 824 人、18 456 户；低收入人口 114 836 人、30 220 户；贫困人口总数 188 660 人。贫困人口数，相当于江南好几个小县的人数总和，远高于国家和省认可的我县 2002 年 625 元以下贫困人口 4.05 万人，625～865 元低收入人口 10.32 万人，也出乎我们一些长期从事贫困监测同志的意料。
- 分布广。我县农村贫困人口分布呈“大集中、小分散”的格局。一方面贫困人口主要集中于五大区域，丘陵山区占一半以上，深圩区次之。另一方面，非贫困乡镇中的非贫困村仍然分布着相当数量的贫困人口，“插花穷”现象比较普遍，即使富甲一方的高沟镇，也不乏纯收入低于 400 元的贫困人口。
- 差距大。从个体看，样本户中 3 000 元以上的高收入户年人均纯收入 3 987.41 元，是绝对贫困户的 9.4 倍，低收入户的 5.5 倍。从总量看，高收入人口与贫困人口大体相等，但前者收入是后者的 7 倍。无为农村楼房林立，但高楼背后有草棚。事实上，正是占全县农户总收入 1/3 的高收入户掩盖了占全县农户总收入 1/22 的贫困群体的存在。这是发展表象背后的贫困实质。
- 程度深。资料显示：在调查的 300 户样本户中，绝对贫困人口人均纯收入仅为 416.41 元，远离贫困线 208.59 元，贫困深度大于全省 20 个重点扶持县的平均数。同时在贫困人口中，又以特困群体居多，近八成的绝对贫困人口纯收入在 500 元以下。可以说，这部分人口是今后扶贫开发工作最难啃的硬骨头，但同时也是最需社会关注和救助的群体。

摘自《扶贫开发简报》第九期 2003.09.12

3.4 离散趋势的描述——变异指标

3.4.1 变异指标的概念

对于一个统计总体或分布，我们计算其平均指标，把总体各单位标志的差异加以概括和抽象，从而反映总体各单位标志值一般水平和集中趋势。但总体内各单位标志值毕竟各不相同，存在着差异，这就须计算反映总体标志值分布的另一数量特征的指标，这就是标志变异指标，它用于测定总体标志值的差异情况和离散程度。在统计研究中，必须把平均指标和变异指标这对相互联系的对应指标结合起来应用，从不同的侧面来揭示总体各单位标志值的分布特征。

3.4.2　变异指标的作用

在统计分析研究中，标志变异指标的作用主要有以下几方面。

1．标志变异指标是衡量平均数代表性大小的尺度

所谓平均数的代表性，是指平均数和总体中各单位标志值的差异情况。平均数代表性好，是指平均数和总体中的各单位标志值差异小，即各标志值的离散程度小；平均数代表性差，是指平均数和总体单位标志值差异大，即各标志值的离散程度大。有三个生产小组工人的日产零件数（个）资料如下：

甲组：80，90，100，110，120

乙组：98，99，100，101，102

丙组：100，100，100，100，100

三组的平均日产量都是 100 个，但各组的离差大小不同，意味着平均数的代表性不同。甲组工人日产量在平均数 100 个周围分布较为分散，各标志值与平均数离差大，平均数代表性低；而乙组工人日产量在平均数周围分布则较为集中，各标志值与平均数离较小，平均数代表性较高；丙组工人日产量均为 100 个，与平均数离差为 0，平均数有完全的代表性。

2．标志变异指标是反映社会经济活动过程均衡性或节奏性的重要指标

在经济发展过程中出现的升降起伏、波动较大的非均衡变化的现象，生产过程中出现的前松后紧或前紧后松的无节奏状况等，可利用标志变异指标对之进行测定和分析。如果标志变异指标值较小，说明现象的发展比较平稳；如果标志变异指标值较大，则表明现象发展中稳定性较差。

3．计算标志变异指标确定推断的准确程度

在抽样调查中，根据样本指标来推断总体指标，只有计算标志变异指标才能确定推断的准确程度及误差大小。

3.4.3　标志变异指标的种类与计算

1．标志变异绝对指标

常用的标志变异绝对指标有极差、平均差、标准差。这一类变异指标主要反映标志变动的绝对程度，用绝对数表示，一般不能用于不同总体之间离散程度大小的直接比较。

（1）极差（R）

极差是指总体或分布中最大标志值与最小标志值的差距，又称全距。用以说明被研究对象中各单位标志值变动的总范围，即

$$R = x_{\max} - x_{\min} \tag{3-24}$$

例如，上述列举的三组工人日产量资料的极差为：

$$R_{甲} = 120 - 80 = 40（个）$$
$$R_{乙} = 102 - 98 = 4（个）$$
$$R_{丙} = 100 - 100 = 0（个）$$

从计算结果可看出，甲组差异最大，乙组其次，而丙组没有差异。R 值越大，说明标志值分布越分散，平均数代表性越差。在分组数列条件下，其极差的近似值为：

$$R = U_{\max} - L_{\min} \tag{3-25}$$

式中，$U_{\max}$ 是最大组的上限；$L_{\min}$ 是最小组的下限。

极差指标计算方法简单、易懂，在实际工作中常用于产品质量的检验和控制。但它只考虑两个极端变量值的水平，不能反映其间的变量分布情况，而且当数列中有异常值存在时，会直接影响极差的大小，就不能确切反映各单位标志值的实际变动范围。

（2）平均差（$A \cdot D$）

平均差是指数列中各单位标志值对其平均数离差绝对值的算术平均数。由于各个标志值对算术平均数的离差总和等于零，所以取离差的绝对值形式计算平均差。

- 未分组资料的简单平均差计算公式为：

$$A \cdot D = \frac{\sum |x - \bar{x}|}{n} \tag{3-26}$$

例如，上例三组工人日产量资料中，计算甲、乙两组的平均差分别为：

$$A \cdot D_{甲} \frac{|80-100| + |90-100| + |100-100| + |110-100| + |120-100|}{5} = 12（个）$$

$$A \cdot D_{乙} = \frac{|98-100| + |99-100| + |100-100| + |101-100| + |102-100|}{5} = 1.2（个）$$

计算表明，乙组的平均差小于甲组，因此乙组平均工资代表性较高。

- 分组数列的加权平均差计算公式为：

$$A \cdot D = \frac{\sum |x - \bar{x}| f}{\sum f} 或 \sum |x - \bar{x}| \frac{f}{\sum f} \tag{3-27}$$

平均差不同于极差，它考虑到总体全部单位标志值的差异，能较准确地反映总体各标志值的平均变异程度，但因需要对离差取绝对值，计算处理过程烦琐，在运用上有较大的局限性，常用的指标是标准差和方差。



➪【例 3-20】

依据表 3-9 资料，计算某公司 50 名员工的月工资水平的平均差。

表 3-9　某公司员工工资平均差计算表

工资 x（元）	员工数 f（人）	xf	$x-\bar{x}$	$\lvert x-\bar{x}\rvert f$	$(x-\bar{x})^2 f$
800	5	4 000	–520	2 600	1 352 000
1 000	10	10 000	–320	3 200	1 024 000
1 200	20	24 000	–120	2 400	288 000
1 500	7	10 500	180	1 260	226 800
2 000	5	10 000	680	3 400	2 312 000
2 500	3	7 500	1 180	3 540	4 177 200
合　计	50	66 000	—	16 400	9 380 000

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{66\,000}{50}=1\,320\ （元）$$

$$A\cdot D=\frac{\sum|x-\bar{x}|f}{\sum f}=\frac{16\,400}{50}=328\ （元）$$

计算结果说明，每个员工的月工资与全部员工的月平均工资平均差为 328 元，在可比的条件下，平均差的数值越大，其平均数的代表性则越小。

（3）标准差（σ）

标准差是总体所有单位标志值与其平均数的离差平方的平均数的平方根，又称均方差。标准差的平方就是方差（σ^2）。标准差计算公式如下。

未分组资料计算简单标准差：

$$\sigma=\sqrt{\frac{\sum(x-\bar{x})^2}{n}} \tag{3-28}$$

分组资料计算加权标准差：

$$\sigma=\sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}} \tag{3-29}$$

标准差的指标分析意义与平均差相似，但指标构造方式却有所不同。标准差和方差的计算过程比平均差简便，数学性质也较为理想，是统计分析中最常用、最重要的变异指标。

➪【例 3-21】

根据例 3-20 的资料，计算该公司 50 名员工的月工资水平的方差和标准差，计算过程如表 3-9 所示。

$$\sigma^2=\frac{\sum(x-\bar{x})^2 f}{\sum f}=\frac{9380\,000}{50}=187\,600\,(元)$$

$$\sigma=\sqrt{\frac{\sum(x-\bar{x})^2 f}{\sum f}}=\sqrt{\frac{9\,380\,000}{50}}=433.13\,(元)$$

2．标志变异相对指标

上面介绍了各种标志变异的绝对指标，这些指标数值大小，不仅受数列平均水平高低的影响，而且都是有名数单位。因此只有在平均数相等单位相同的条件下，才适宜采用这类指标直接比较不同总体变量值的离散程度。为了对比分析不同平均水平的总体的变异程度，必须消除平均数和计量单位不同的影响，就需要计算标志变异相对指标。

标志变异相对指标是和各类标志变异的绝对指标与相应的平均数之比来表明标志变异的相对程度，又称为标志变异系数或离散系数，用 V 表示。它可以消除数列平均水平高低对标志变异程度大小的影响，反映不同水平和不同性质的变量数列的变异程度。

（1）全距系数（V_R）

$$V_R=\frac{R}{\bar{x}}\times 100\% \qquad (3\text{-}30)$$

（2）平均差系数（$V_{A\cdot D}$）

$$V_{A\cdot D}=\frac{A\cdot D}{\bar{x}}\times 100\% \qquad (3\text{-}31)$$

（3）标准差系数（V_σ）

$$V_\sigma=\frac{\sigma}{\bar{x}}\times 100\% \qquad (3\text{-}32)$$

其中，标准差系数在实际应用中最广泛。

➪【例 3-22】

现以表 3-10 两个企业的工人劳动生产率及标准差资料，说明标志变异系数的计算与分析。

表 3-10　两个企业的标准差系数计算表

企业	工人平均劳动生产率 $\bar{x}$（元）	标准差 σ（元）	标准差系数 V_σ（%）
甲	16 000	600	3.75
乙	8 000	400	5.00

从表 3-10 中可知，虽然甲企业标准差大于乙企业，但不能由此说明甲企业工人平均劳动生产率的代表性比乙企业小，这是因为两个企业的工人平均劳动生产率水平不同，所以不能只根据标准差大小作结论。为了说明两企业的变异程度，只有通过标志变异系数才能比较，因为它消除了不同数列平均水平的影响。例中计算标准差系数后表明，甲企业的标志变异程度比乙企业小，因而正确分析的结论是甲企业的工人平均劳动生产率更具有代表性。

本章小结

1. 平均指标是反映总体一般水平的代表值和描述变量分布集中趋势的重要特征值，在社会经济统计中常用的平均指标有算术平均数、调和平均数、几何平均数、众数、中位数等。

2. 各种平均指标都有其特点和适用的条件，它们的计算方法不同，对同一资料的计算结果也不同。因此，必须根据研究的具体目的和数据分布特点，正确选择平均指标。在社会经济统计中，计算和运用平均指标还必须注意：只有同质的量才能平均；平均数与典型值和变量分布相结合；总平均数与组平均数相结合；平均数与变异指标相结合。

3. 变异指标反映分布的离散趋势，是与平均指标相匹配的重要特征值。常用的变异指标有全距、四分位差、平均差、标准差（或方差）等。其中，方差和标准差是最重要、应用最广泛的一种。

4. 不同总体的平均差、标准差和方差不仅因计量单位不同而不可比，当它们的平均数相差很大时也不可比。对两个总体的变异程度进行比较时，为了消除这种不可比因素，就要采用标志变异的相对数指标——离散系数。最常用的离散系数是平均差系数和标准差系数。

复习思考题

一、概念题

简单算术平均数　加权算术平均数　调和平均数　几何平均数　众数　中位数　全距　平均差　标准差　离散系数　标准差系数

二、简答题

1. 时期总量指标与时点总量指标有何异同?

2. 强度相对指标与平均指标有何区别?
3. 在分析长期计划执行情况时，水平法与累计法有何区别?
4. 在统计实践中，为什么要将各种指标结合起来使用?
5. 什么是平均指标? 其特征如何? 作用如何?
6. 什么是加权算术平均数? 加权是何意思? 权数对平均数起什么作用?
7. 在计算平均指标时，算术平均数、调和平均数、几何平均数分别适用于什么条件?
8. 什么是众数和中位数? 如何确定?
9. 如何运用算术平均数、众数、中位数反映分布的特征?
10. 什么是标志变异指标? 它有哪些种类?
11. 简单说明平均指标与变异指标在反映总体特征方面的联系与区别?
12. 什么是标志变异相对指标? 它在什么情况下使用?

三、练习题

1．判断题（把“√”或“×”填在题后的括号里）

（1）算术平均数的大小，只受总体各单位标志值大小的影响。（　　）

（2）在特定条件下，加权平均数等于简单算术平均数。（　　）

（3）标志变异指标数值越大，说明总体中各单位标志值的差异程度就大，则平均数的代表性就小。（　　）

（4）任何两个总体，比较其平均数的代表性大小时，都可以采用标准差比较。（　　）

（5）平均差是总体各单位标志值与其算术平均数离差平方的算术平均数的平方根。（　　）

2．单选题

（1）在变量数列中，各组次数（权数）均扩大100倍，则算术平均数就（　　）。

A. 扩大100倍　B. 缩小100倍　C. 数值不变　D. 不能确定

（2）已知某市的甲、乙、丙三个农贸市场的某种蔬菜的单价和购买额，计算这种蔬菜的平均价格应采用（　　）。

A. 简单算术平均数　B. 加权算术平均数

C. 简单调和平均数　D. 加权调和平均数

（3）标准差数值越小，则反映各标志值（　　）。

A. 越分散，平均数的代表性越小　B. 越集中，平均数的代表性越大

C. 越分散，平均数的代表性越大　D. 越集中，平均数的代表性越小

（4）在各标志变异指标中，最易受极端影响的是（　　）。

A. 极差　B. 平均数　C. 标准差　D. 标准差系数

（5）为了用标准差比较两个总体的平均数的代表性，其基本条件是（　　）。

A. 两个总体的单位数相等　B. 两个总体的平均数相等

C. 两个总体的平均差相等　　D. 两个总体的离差之和应相等

（6）两组工人生产同样零件，甲组工人每人加工零件数为：32 件，25 件，29 件，28 件，26 件；乙组为：30 件，25 件，22 件，36 件，27 件。这两组工人加工零件数的差异程度（　　）。

A. 甲组大于乙组　　B. 乙组大于甲组

C. 两组相同　　D. 无法比较

（7）两个总体的平均数不等，但标准差相等，则（　　）。

A. 平均数小，代表性大　　B. 平均数大，代表性大

C. 代表性相同　　D. 无法正确判断

（8）已知两个同类型企业，职工平均工资的标准差分别为 7 元和 6 元，则两个企业职工平均工资的代表性是（　　）。

A. 甲大于乙　　B. 乙大于甲　　C. 一样的　　D. 无法判断

3．多选题

（1）下列指标属于强度相对指标的有（　　）。

A. 每百户居民拥有电话机的数量　　B. 人均粮食产量

C. 人均钢产量　　D. 人均生活费支出

E. 粮食平均亩产量

（2）某小组三名工人日产零件数分别为 120 件、104 件、190 件，根据这一资料得出（　　）。

A. 全距大于平均差　　B. 全距大于标准差

C. 标准差大于平均差　　D. 标准差大于标准差系数

4．计算题

（1）某企业 2008 年的劳动生产率计划规定比上年提高 8%，实际执行结果比上年提高 10%，问劳动生产率计划完成程度是多少?

（2）某厂按计划规定，第一季度的单位产品成本比去年同期降低 10%，实际执行结果比去年同期降低 8%，问该厂第一季度单位产品成本计划完成程度如何?

（3）某市“五年计划”规定，计划期最后一年产品产量应达到 70 万吨，实际情况如表 3-11 所示。

表 3-11　某市“五年计划”情况　　单位：万元

时间	第一年	第二年	第三年		第四年				第五年			
			上半年	下半年	第一季度	第二季度	第三季度	第四季度	第一季度	第二季度	第三季度	第四季度
产量	45	48	25	27	16	16	18	17	18	20	23	25

试计算该市甲产品产量五年计划完成程度和提前完成的时间。

（4）某城市某年土地面积和人口资料如表 3-12 所示。试计算有关的强度相对指标、比较相对指标、比例相对指标、结构相对指标，并作简要分析。

表 3-12　土地面积和人口资料

	土地面积（平方公里）	年平均人口（万人）
市区	587.4	741.3
郊区	1 432.6	534.2
合计	2 020.0	1 275.5

（5）某厂 400 名职工工资资料如表 3-13 所示。试计算该厂职工平均工资和标准差。

表 3-13　某厂职工工资资料

按月工资分组（元）	职工人数（人）
450～550	60
550～650	100
650～750	140
750～850	60
850～950	40
合　　计	400

（6）甲、乙两个农贸市场三种主要蔬菜价格及销售额资料如表 3-14 所示。试比较该地区哪个农贸市场蔬菜的平均价格高，并说明其原因。

表 3-14　甲、乙两个农贸市场三种蔬菜的情况

品种	价格（元/千克）	销售额（万元）	
		甲市场	乙市场
A	0.30	75.0	37.5
B	0.32	40.0	80.0
C	0.36	45.0	45.0

（7）某投资银行在 10 年中的年利率是：有 1 年为 7%，有 3 年为 8%，有 4 年为 10%，有 2 年为 11%，若按复利计算，试求这 10 年的年平均利率是多少?

（8）某地区抽样调查职工家庭收入资料如表 3-15 所示。试计算职工家庭平均每人月收入的众数、中位数和算术平均数，并简单说明其分布特征。

表 3-15　某地区职工家庭收入情况

按平均每人月收入分组（元）	职工户数（户）
700 以下	40
700～750	100
750～800	170
800～850	220
850～900	190
900～950	150
950～1 000	130
1 000 以上	120
合　计	1 120

（9）某地科学试验站对 A、B 两个品种的水稻分别在 4 块地进行试验，其产量如表 3-16 所示。试根据表 3-16 资料分别计算两个品种的平均单位面积产量，并确定哪一个品种具有较大的稳定性。

表 3-16　A、B 两个品种水稻的产量

	A　品　种		B　品　种	
序号	田地面积（公顷）	产　量（千克）	田地面积（公顷）	产　量（千克）
1	0.08	600	0.07	497
2	0.05	405	0.09	675
3	0.10	725	0.05	375
4	0.09	720	0.10	700

CHAPTER 4

第 4 章　抽样估计

章首引例

茶叶包装过程的检验

在龙津茶叶公司的生产包装车间，每 8 小时的工作时间要完成 5 000 多袋茶叶的装袋任务。作为工厂的品管经理，要负责监督每袋茶叶的包装重量和质量。平均来说，每袋茶叶的等级是袋子所标示的等级，重量是 100 克。然而，受装袋过程速度的影响，每袋茶叶的实际重量都不相同，甚至出现等级与标示的差异。如果明显高于 100 克，则将导致产品成本上升，公司利益受损；如果明显低于 100 克，则将损害消费者利益，并将受到经销商退货甚至有关部门查处，公司声誉和形象受损。由于商品逐袋检验费时费力、效率低下，必须采用抽样调查的方法，并根据抽查样本结果来判断装袋过程正常与否。品管人员采用每次抽取一个样本对其称重和等级核对，然后计算出样本的平均重量 $\bar{x}$ 和等级合格率 p 。然后判断这个 $\bar{x}$ 取自于一个均值 μ 为 100 克的总体的可能性（或者总体等级合格率为p的可能性），据此做出是保持还是改变甚至关停袋装过程的决定。[①]

本章学习目标

- ☑ 理解抽样估计的概念、特点、作用以及几个基本概念；
- ☑ 掌握抽样误差的含义和影响抽样误差的主要因素；
- ☑ 熟练掌握抽样平均误差的计算；
- ☑ 熟练掌握抽样估计的两种方法：点估计和区间估计；
- ☑ 掌握必要抽样数目的确定方法。

① 摘自孙允午主编的《统计学》，上海财经大学出版社，2006。

4.1 抽样估计的一般问题

4.1.1 抽样估计的意义与特点

抽样估计又称抽样推断，它是在抽样调查的基础上，利用样本的实际资料推断总体相应数量特征的一种统计分析方法。在实际工作中，许多场合我们没有可能或不必要对总体的所有单位进行全面调查，来达到对总体数量特征的认识。

- 对于无限总体，统计上无法进行全面调查了解，只有借助于抽样推断的方法来认识总体的数量特征。例如，在连续大量生产的某种小件产品（如牙签、螺母等）中，总产量是无限的，只能进行抽样推断。
- 许多产品的例行质量检验是带有破坏性的或消耗性的。例如，灯泡寿命试验要一直长期点亮直到烧毁，这是破坏性试验；烟、酒的质量品尝均属消耗性质量检验。对这些总体都无法进行全面调查，只能进行抽样调查。
- 对于某些现象，虽然可以进行全面调查，但需要花费大量的人力、物力、财力和时间，若采用抽样调查，可以达到事半功倍的效果。例如，要了解水库中的鱼苗数、森林的木材积蓄量、居民对主要耐用消费品的需求量等，适宜采用抽样调查进行推断。
- 对全面调查的资料进行评价与修正。例如，我国人口普查规定，在人口普查工作完毕后，还要按照规定的抽样方法抽取若干地区的人口进行复查。用抽样调查的资料，计算人口全面调查的差错率。再根据这个比率去修正普查资料，从而保证人口调查资料的质量，使调查资料更为准确，更接近于实际的数值。
- 在对工业生产过程的质量控制。对于成批或大量连续生产的产品生产过程，通过抽样方法可以及时提供有关产品的质量信息，分析各种可能的原因，以便采取措施，排除故障，使生产过程保持正常，从而起到对生产过程进行质量控制的作用。

在上述几种情况中，都很难对每个单位进行观察，只能组织抽样调查，取得部分的实际资料，来估计和推断总体的数量特征，以达到对现象总体的认识。

抽样估计有如下特点。

1．抽样调查是由部分推算整体的一种认识方法

抽样调查是一种非全面调查，但调查的目的却不在于了解部分单位的情况，它只是作为进一步推断的手段，目的乃在于要认识总体的数量特征。那么根据抽样调查资料能否推断出总体的特征呢？例如，检测几克棉花纤维的强度能不能判断整批棉花纤维的强度？又如，对几颗种子进行催芽试验能不能判断该品种整批种子的发芽率；等等。如果在方法上不能解决这类问题，那么统计的认识活动就要受到限制，统计科学也很难得到发展。抽样推断原理解决了这一问题，它科学地论证了样本指标与相应的总体指标之间存在着内在的

联系，两者误差的分布也是有规律可循的，并提供一套利用抽样调查的部分信息来推断总体数量特征的方法，这就大大提高了统计分析的认识能力，为信息采集和开发开辟了一条崭新的途径。

2．抽样估计是建立在随机取样的基础上的

随机原则就是总体中样本单位的中选或不中选，不受主观因素的影响，保证每一单位都有相等的中选可能性。把抽样估计建立在随机样本的基础上，才可能事先掌握各种样本出现的可能性大小，提供样本指标数值的分布情况，计算样本指标的抽样平均误差，同时估计样本指标与总体指标的抽样误差不超过一定范围的概率保证程度。只有坚持抽样的随机原则，抽样估计才可能利用概率论原理来研究样本指标与总体指标的关系，确定优良估计的标准，为抽样设计寻求更有效的抽样组织形式建立科学的理论基础。

3．抽样估计是运用概率估计的方法

利用样本指标来估计总体指标，在数学上运用不确定的概率估计法，而不是运用确定的数学分析法。抽样估计原则上把由样本观察值所决定的样本指标看做随机变量。在实践中抽取一个样本，计算样本指标值并将其作为相应总体指标的估计值。接着需要研究的问题是，用这样的样本指标值来代表相应的总体指标值的可靠程度究竟有多大，即概率估计所要解决的问题。例如，要知道全市职工的人均收入，首先抽取若干企业作为样本，计算样本的职工人均收入为 750 元，又求得以 750 元来代表全市职工的人均收入的误差不超过 50 元的概率不低于 90%。如果这一估计的可靠程度被认为已经满足分析工作的要求，那么可以用 750 元作为全市职工的人均收入水平，否则就要改善抽样组织，重新进行抽样调查，以提高结论的可靠程度，这便是概率估计的基本思路。

4．抽样估计的误差可以事先计算并加以控制

以样本指标估计相应的总体指标虽然也存在一定的误差，但它与其他统计估算不同。抽样误差范围可以事先通过有关资料加以计算，并且可以采取必要的组织措施来控制这个误差范围，保证抽样推断的结果达到一定的可靠程度。也可以这样说，抽样调查就是根据事先给定的误差允许范围进行设计的，而抽样估计则是具有一定可靠程度的估计和判断，这些都是其他估算方法办不到的。

4.1.2 有关抽样的基本概念

1．总体和样本

总体指所要认识的研究对象全体，它是由所研究范围内具有某种共同性质的全体单位组成的集合体。总体的单位数通常都是很大的，甚至是无限的，这样才有必要组织抽样调查。一般用 N 来表示总体的单位数。在组织抽样调查时首先要弄清总体的范围、单位的含

义，以及可实施的条件，以清单、名册、图表等形式编制抽样框作为抽样的母体。

样本是从总体中随机抽取出来的，作为代表这一总体的部分单位组成的集合体。样本的单位数总是有限的，相对总体来说它的数目比较小，一般用 n 来表示样本的单位数。一般来说，样本单位数达到或超过 30 个称为大样本，30 个以下称为小样本。社会经济现象的抽样调查多取大样本。

作为估计对象的总体是确定的，而且是唯一的。但作为观察对象的样本就不是这样。从一个总体中可以抽取很多个样本，每次可能抽到哪个样本不是确定的，也不是唯一的，而是可变的。明白这一点对于理解抽样推断原理是很重要的。

2．总体指标和样本指标

根据总体各单位的标志值或标志属性计算的，反映总体数量特征的综合指标称为总体指标，也称为参数。由于总体是唯一确定的，所以根据总体计算的总体指标也是确定的、唯一的。对于总体中的数量标志，常用的总体指标有总体平均数 $\bar{X}$ 、总体方差 σ^2（或总体标准差 σ）。

设总体变量 X 为：X_1 ，X_2，… ，X_n，则有：

$$\bar{X}=\frac{\sum X}{N}\text{ 或 }\frac{\sum XF}{\sum F},\quad s^2=\frac{\sum(X-\bar{X})^2}{N}\text{ 或 }\frac{\sum(X-\bar{X})^2F}{F}$$

对于总体中的品质标志，由于各单位的标志不能用数量来表示，因此总体指标常以成数指标 P 来表示总体中具有某种性质的单位数在总体全部单位数中所占的比重，以 Q 表示总体中不具有某种性质的单位数在总体中所占的比重。

设总体 N 个单位中，有 N_1 个单位具有某种性质，N_0 个单位不具有某种性质，$N_1+N_0=N$，则有：

$$P=\frac{N_1}{N}\text{，}\quad Q=\frac{N_0}{N}=\frac{N-N_1}{N}=1-P$$

品质标志的表现只有“是”、“非”两种。例如，产品质量标志表现为合格品和不合格品。可以把“是”的标志表现表示为 1，而把“非”的标志表现表示为 0。那么成数 P 就可以视为（0，1）分布的平均数，并可以求相应的方差和标准差，公式如下：

$$\bar{X}_P=\frac{\sum XF}{\sum F}=\frac{0N_0+1N_1}{N_0+N_1}=\frac{N_1}{N}=P(1-P)$$

$$\sigma_P^2=\frac{(0-P)^2N_0+(1-P)^2N_1}{N_0+N_1}=\frac{P^2N_0+Q^2N_1}{N}=P^2Q+Q^2P=PQ(P+Q)=P(1-P)$$

例如，某批零件的合格品率 $P=80\%$，则有

$$\bar{X}_P=80\%\text{，}\quad \sigma_P^2=80\%\times 20\%=16\%$$

在抽样估计中，总体指标的意义和计算方法是明确的，但指标的具体数值事先是未知的，需要用抽样来估计它。

根据样本各单位标志值或标志属性计算的综合指标为样本指标或抽样指标。样本指标是样本变量的函数，用来估计总体指标，因此和常用的总体指标相对应，有样本平均数、样本方差和样本成数等。一般以小写字母表示。

设样本变量 x 为 $x_1, x_2, \cdots, x_n$，则有：

$$\bar{x} = \frac{\sum x}{n} \text{或} \frac{\sum xf}{\sum f}, s_x^2 = \frac{\sum (x - \bar{x})^2}{n} \text{或} \frac{\sum (x - \bar{x})^2 f}{\sum f}$$

$$\bar{x}_p = \frac{n_1}{n} = p, s_p^2 = p(1 - p)$$

样本指标的计算方法是确定的，但由于一个总体可以抽取多个样本，样本不同，样本指标的数值也就不同，所以抽样指标的数值不是唯一确定的。抽样指标是样本变量的函数，是随机变量。所以把它作为总体指标的估计值，有的误差大些，有的误差小些，有的发生正误差，有的发生负误差，情况各不相同。

3．重复抽样和不重复抽样

从抽样的方法来看，抽样可以有重复抽样和不重复抽样两种。

（1）重复抽样，也称回置抽样

它是这样安排的，要从总体 N 个单位中随机抽取一个容量为 n 的样本，每次从总体中抽取一个单位，把结果登记下来，又重新放回，参加下一次抽选。因而重复抽样的样本是由 n 次相互独立的抽选构成的，每次抽选是在完全相同的条件下进行的，每个单位中选的机会在各次都完全相等。

从总体 N 个单位中，用重复抽样的方法，随机抽取 n 个单位构成一个样本，则共可抽取 N^n 个样本。

例如，总体有A，B，C，D四个单位，要从中以重复抽样的方法抽取2个单位构成样本。先从4个单位中取1个，共有4种取法。结果登记后再放回，然后再从相同的4个中取1个，也有4种取法。前后取两个构成一个样本，全部可能抽取的样本数目为 $4 \times 4 = 16$ 个，它们是：AA，AB，AC，AD，BA，BB，BC，BD，CA，CB，CC，CD，DA，DB，DC，DD。

（2）不重复抽样，也称为不回置抽样

它是这样安排的，要从总体 N 个单位中抽取一个容量为 n 的样本，每次从总体中抽取一个单位，连续进行 n 次抽样构成一个样本，但每次抽取一个单位就不再放回参加下一次的抽选。因而不重复抽样有这样的特点：样本由 n 次连续抽取的结果构成，实质上等于一次同时从总体中抽取 n 个样本单位，连续 n 次抽选的结果不是相互独立的，每次抽取的结果都影响下一次抽取，每抽取一次总体单位数就少一个，因而每个单位的中选机会在各次是不相同的。

从总体 N 个单位中，用不重复抽样的方法，抽取 n 个单位样本，全部可能抽取的样本数目为 $N(N-1)(N-2)\cdots(N-n+1)$ 个。

例如，从 A，B，C，D 四个单位中，用不重复抽样的方法从中抽取 2 个单位构成样本。先从 4 个单位中取 1 个，共有 4 种取法，第二次再从留下的 3 个单位中取 1 个，共有 3 种取法，前后两个构成一个样本，全部可能抽取的样本数目为 $4\times3=12$ 个，它们是：

AB，AC，AD，BA，BC，BD，CA，CB，CD，DA，DB，DC。

可见在相同样本容量的要求下，重复抽样的样本个数总是大于不重复抽样的样本个数。

4.2 抽样误差

4.2.1 抽样误差的含义

用抽样指标来估计全及指标是否可行，关键在于抽样误差。抽样误差大小表明抽样效果好坏，如果误差超过了允许的限度，抽样调查也就失去了价值。

调查误差是指调查所获得的统计数据与调查总体未知真实数据之间的差别，它包括登记性误差和代表性误差两种。

（1）登记性误差

它是在调查过程中由主客观原因而引起的登记性差错造成的误差。比如，由测量、记录、计算造成的差错，由被调查者所报材料不实或调查者有意虚报、瞒报资料造成的差错，等等。对于这类误差可以通过搞好调查的宣传组织工作、不断提高调查人员素质、广泛采用电子计算机技术等措施加以解决。

（2）代表性误差

它是用样本指标数值去推算总体指标数值时，由于样本结构情况不足以代表总体特征所产生的误差。这类误差又可分为两种：一种是由于没有遵守随机原则而造成的误差，通常称为系统性误差。例如，调查者有意挑选较好或较差的单位作为样本；另一种误差是即使遵守了随机原则，也会由于被抽取的样本各种各样，导致样本内部各单位的分布比例结构与总体实际分布状况有偶然性的差异，从而使不同随机样本得出不同的估计量，造成样本指标数值与总体指标数值之间产生差距，如抽样平均数与总体平均数的离差 $(\bar{x}-\overline{X})$ 、抽样成数与总体成数的离差（$p-P$）等。这类误差通常称为抽样误差或随机误差。可见，只要遵循随机原则，系统性误差是可以消除的。但是，用样本估算总体，两者之间总是要出现差距的。抽样误差是抽样法本身所固有的误差，是不可避免的，也无法消除，但是可以控制。

4.2.2 影响抽样误差的主要因素

1．样本单位数的多少

在其他条件不变的情况下，抽样单位数越多就越能反映总体，误差就越小。当抽样单

位数接近总体单位数时，此时的抽样调查已接近于全面调查，抽样误差接近于零；反之，抽样单位数越少，误差越大。

2．总体被研究标志变异程度的大小

在其他条件不变的情况下，所研究总体的标志变异程度越小，说明总体各单位标志值之间的差异越小，这样抽样指标与总体指标之间的误差也越小。相反，若总体被研究标志变异程度大，则抽样指标与总体指标之间的误差也大。

3．抽样组织方式

采取不同的抽样组织方式，所抽出的样本对于总体的代表性也不相同，因此抽样组织方式影响抽样误差的大小。在实践中，我们可以利用不同抽样组织方式下抽样误差的大小来判断不同方式的有效性。

4．抽样方法

抽样方法有重复抽样和不重复抽样两种。在其他条件相同时，不重复抽样的抽样误差一般小于重复抽样的误差，这是因为不重复抽样避免了总体单位的重复中选，因而更能反映总体结构，故抽样误差会较小些。

4.2.3 抽样平均误差

在一定组织方式下搞一次抽样调查，根据统计研究的目的和任务，可以取一个样本，也可以取多个样本。在抽样多个样本时，就其中每个样本来说，都有其相应的样本指标。

由于样本是按随机原则抽取的，所以在同一总体中，按相同的抽样数目，可以抽出多个相同和不同的样本，而每次抽出的样本都可以计算出相应的抽样平均数、抽样成数和抽样误差。从理论上说，可以计算出多个抽样误差，它们带有偶然性，有的可能是正误差，有的可能是负误差；有的可能性大一些，有的可能性小一些。为了用样本指标去推算总体指标，需要计算这些抽样误差的平均数，即抽样平均误差，用来衡量抽样误差的一般水平。

通常用抽样平均数（或成数）的标准差作为衡量其抽样误差一般水平的尺度。按照标准差的一般定义，抽样平均数（或成数）的标准差是按抽样平均数（或成数）与其平均数的离差平方和计算的，但由于抽样平均数的平均数等于总体平均数，而抽样成数的平均数等于总体成数，抽样指标的标准差恰好反映了抽样指标和总体指标的平均离差程度。

抽样平均误差以符号 μ 表示，以 $\mu_{\bar{x}}$ 代表平均数的抽样平均误差，以 μ_p 代表成数的抽样平均误差，以 K 代表可能组成的样本总数。

1．计算抽样平均误差的理论公式

根据抽样平均误差的概念可得其一般计算公式：

$$\mu=\sqrt{\frac{\sum(\text{样本指标}-\text{总体指标})^2}{\text{可能组成的样本总数}}}$$

即

$$\mu_{\bar{x}}=\sqrt{\frac{\sum(\bar{x}-\bar{X})^2}{K}} \tag{4-1}$$

$$\mu_p=\sqrt{\frac{\sum(p-P)^2}{K}} \tag{4-2}$$

➪【例 4-1】

假设有 4 个工人，其每日产量为 70，90，130，150 件，从中随机抽取 2 个工人作为样本，求抽样平均误差。

若采取重复抽样法，可能配合的样本数目及相应指标的计算见表 4-1。

表 4-1　重复抽样的抽样平均误差计算表

序号	样本变量（x）	样本平均数（$\bar{x}$）	平均数离差（$\bar{x}-\bar{X}$）	离差平方 $(\bar{x}-\bar{X})^2$
1	70，70	70	－40	1 600
2	70，90	80	－30	900
3	70，130	100	－10	100
4	70，150	110	0	0
5	90，70	80	－30	900
6	90，90	90	－20	400
7	90，130	110	0	0
8	90，150	120	10	100
9	130，70	100	－10	100
10	130，90	110	0	0
11	130，130	130	20	400
12	130，150	140	30	900
13	150，70	110	0	0
14	150，90	120	10	100
15	150，130	140	30	900
16	150，150	150	40	1 600
合计	—	1 760	0	8 000

全部可能的样本平均数的平均数（$\bar{x}$）会等于总体平均数（$\bar{X}$），故得：

$$\bar{x}=\frac{\sum x_i}{K}=\frac{1\ 760}{16}=110(\text{件})=\bar{X}$$

全部可能组成的样本标准差为：

$$\mu_{\bar{x}}=\sqrt{\frac{\sum(\bar{x}_i-\bar{X})^2}{K}}=\sqrt{\frac{8\,000}{16}}=22.36(\text{件})$$

22.36 件是 16 个可能的样本平均数的标准差，称为抽样平均误差（μ）。

若采用不重复抽样法，可能的样本数及相应指标的计算如表 4-2 所示。

表 4-2　不重复抽样的抽样平均误差计算表

序号	样本变量（x）	样本平均数（$\bar{x}$）	平均数离差（$\bar{x}-\bar{X}$）	离差平方 $(\bar{x}-\bar{X})^2$
1	70，90	80	–30	900
2	70，130	100	–10	100
3	70，150	110	0	0
4	90，70	80	–30	900
5	90，130	110	0	0
6	90，150	120	10	100
7	130，70	100	–10	100
8	130，90	110	0	0
9	130，150	140	30	900
10	150，70	110	0	0
11	150，90	120	10	100
12	150，130	140	30	900
合计	—	1 320	0	4 000

$$\mu=\sqrt{\frac{\sum(\bar{x}_i-\bar{X})^2}{K}}=\sqrt{\frac{4\,000}{12}}=18.36\ (\text{件})$$

18.26 件是 12 个可能配合的样本平均数的标准差，即抽样平均误差。它比重复抽样的平均误差 22.36 小。

需要指出，式（4-1）和式（4-2）表明了抽样误差的意义，但在实际中总体的 $\bar{X}$ 和 P 是未知的，而且也无法计算全部样本的平均数和成数，所以按上述公式计算抽样平均误差实



际上是不可能的。在实践中可以通过其他方法加以推算。

2．抽样平均误差的计算公式

数理统计证明，在纯随机抽样方式下，抽样平均误差可以采用下面的公式计算。

（1）抽样平均数的平均误差

- 在重复抽样条件下：

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}} \tag{4-3}$$

- 在不重复抽样条件下：

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)} \tag{4-4}$$

当 N 很大时，

$$\mu_{\bar{x}}=\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} \tag{4-5}$$

（2）抽样成数的平均误差

- 在重复抽样条件下：

$$\mu_p=\sqrt{\frac{P(1-P)}{n}} \tag{4-6}$$

- 在不重复抽样条件下：

$$\mu_p=\sqrt{\frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)} \tag{4-7}$$

当 N 很大时，

$$\mu_p=\sqrt{\frac{P(1-P)}{n}\left(1-\frac{n}{N}\right)} \tag{4-8}$$

根据例 4-1 资料，计算总体平均数、标准差如下：

$$\overline{X}=\frac{\sum X}{N}=\frac{440}{4}=110(\text{件})$$

$$\sigma=\sqrt{\frac{\sum(x-\overline{X})^2}{N}}=\sqrt{\frac{(70-110)^2+(90-110)^2+(130-110)^2+(150-110)^2}{4}}$$

$$=31.62(\text{元})$$

现在直接按式（4-3）、式（4-4）计算抽样平均误差 $\mu_{\bar{x}}$。

在重复抽样条件下：

$$\mu_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{31.62}{\sqrt{2}} = 22.36 \text{ (元)}$$

在不重复抽样条件下：

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})} = \sqrt{\frac{31.62^2}{2}(\frac{4-2}{4-1})} = 18.26(\text{元})$$

两者计算结果完全相同。

应用上述公式计算抽样平均误差时要注意两点。

第一，上式中的标准差σ和成数 P 是总体的标准差和成数，而总体的指标通常是未知的，一般用经验数据或样本的标准差 s 和成数 p 来代替，得到近似值。

样本标准差 s 计算公式是：$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ 式中，x 为样本变量；$\bar{x}$ 为样本平均数；$n-1$ 为样本变量自由度。因为 s 是在估计 $\bar{x}$ 的基础上进行第二次估计，所以失去一个自由度。用这一公式计算，才是总体σ^2的无偏估计。但在大样本的情况下，也可以用样本的标准差一般公式即以 n 为分母计算，来代替总体标准差。

第二，上面不重复抽样公式中，如果抽样单位数相对较少，而总体单位数相对很多，则（$1-n/N$）这个系数接近于 1，乘上这个系数后，对平均误差的影响不大。为了简化，在实际工作中，对不重复抽样的情况也往往采用重复抽样公式计算抽样平均误差。

➪【例 4-2】

估计某地区 10 000 名适龄儿童的入学率。随机从这一地区抽取 400 名儿童，检查有 320 名儿童入学，求抽样入学率的平均误差。

根据已知条件：

$$p = \frac{320}{400} = 80\%,\ s^2 = p(1-p) = 80\% \times 20\% = 16\%$$

- 在重复抽样的情况下，入学率的抽样平均误差为：

$$\mu_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.16}{400}} = 2\%$$

- 在不重复抽样的情况下，入学率的抽样平均误差为：

$$\mu_p = \sqrt{\frac{p(1-p)}{n}(1-\frac{n}{N})} = \sqrt{\frac{0.16}{400}(1-\frac{400}{10\ 000})} = 1.96\%$$

计算结果表明，用样本的入学率估计总体的入学率的误差的绝对值在 2%左右。

4.2.4　抽样极限误差

抽样极限误差是从另一角度考虑抽样误差问题。以样本的抽样指标来估计总体指标，要达到完全准确毫无误差，这几乎是不可能的事情。所以在估计总体指标的同时就必须估计误差的大小。我们不希望误差太大，因为误差越大，样本的价值便越小。但也不是误差越小越好，因为在一定限度之外减少抽样误差势必增加很多费用。所以在做抽样估计时，应该根据所研究对象的变异程度和分析任务的要求确定可允许的误差范围，在这个范围内的数字都算是有效的。我们把这种可允许的误差范围称为抽样极限误差。它等于样本指标可允许变动的上限或下限与总体指标之差的绝对值。

设 $\Delta_{\bar{x}}, \Delta_p$ 分别表示抽样平均数极限误差和抽样成数极限误差。则有：

$$\Delta_{\bar{x}} = \left|\bar{x} - \bar{X}\right|, \quad \Delta_p = \left|p - P\right|$$

上式 $\bar{x}$ 与 p 都表示样本平均数和样本成数可允许的上限或下限数值。将上面的等式变换为下列等价的不等式关系：

$$\bar{x} - \Delta_{\bar{x}} \leqslant \bar{X} \leqslant \bar{x} + \Delta_{\bar{x}}$$
$$p - \Delta_p \leqslant P \leqslant p + \Delta_p$$

上面第一不等式表示被估计的总体平均数以抽样平均数 $\bar{x}$ 为中心；被包含在 $\bar{x} - \Delta_{\bar{x}}$ 至 $\bar{x} + \Delta_{\bar{x}}$ 之间，区间［ $\bar{x} - \Delta_{\bar{x}}$ ， $\bar{x} + \Delta_{\bar{x}}$ ］称为平均数的估计区间或称平均数的置信区间，区间的总长度为 $2\Delta_{\bar{x}}$，在这个区间内样本平均数和总体平均数之间的绝对离差不超过 $\Delta_{\bar{x}}$。同样，上面第二不等式表明被估计的总体成数 P 被包含在以抽样成数 p 为中心的 $p - \Delta_p$ 至 $p + \Delta_p$ 的区间内。在［ $p - \Delta_P$ ， $p + \Delta_P$ ］区间内，抽样成数与总体成数之间的绝对离差不超过 $2\Delta_p$。

⇨【例 4-3】

要估计某市居民每月平均用水量和用水量在 10 吨以上的居民户占全部居民的比例，从全市居民中抽取 400 户，得到户均用水量为 5 吨，如果确定抽样极限误差为 0.08 吨，这就要求户均月用水量是 5 ± 0.08 吨，即在 $4.92 \sim 5.08$ 吨之间；由于户均用水量在 10 吨以上的占 30%，又确定抽样极限差为 3%，这就要求居民中，户均用水量在 10 吨以上的比重为 $30\% \pm 3\%$之间，即在 $27\% \sim 33\%$之间。

4.2.5　抽样误差的概率度

根据概率估计的要求，抽样极限误差是以抽样平均误差 $\mu_{\bar{x}}$ 或 m_p 为标准单位来衡量的。

把极限误差 $\Delta_{\bar{x}}$ 或 Δ_p 分别除以 $\mu_{\bar{x}}$ 或 m_p 得相对数 z，t 表示误差范围为抽样平均误差的倍数，是测量估计可靠程度的一个参数，称为抽样误差的概率度。

$$t=\frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}},\quad \Delta_{\bar{x}}=t\mu_{\bar{x}} \tag{4-9}$$

$$t=\frac{\Delta_p}{\mu_p},\quad \Delta_p=t\mu_p \tag{4-10}$$

如上例中，若又已知该市户均用水量的标准差为 0.8 吨，n=400 户，则按重复抽样计算抽样平均误差为：

$$\mu_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{0.8}{\sqrt{400}}=0.04(\text{吨})$$

我们就可以用概率度 $t=\Delta_{\bar{x}}/\mu_{\bar{x}}=0.08/0.04=2$ 来表示极限误差的范围，即以 $2\mu_{\bar{x}}$ 来规定误差范围的大小，这时就要求该市居民户均月用水量在 $5\pm2\mu_{\bar{x}}$ 之间。

再如，已知 400 户居民中月均用水量在 10 吨以上的户数可占比重为 30%，则可求得其抽样平均误差为：

$$\mu_p=\sqrt{\frac{p(1-p)}{n}}=\sqrt{\frac{30\%\times70\%}{400}}=2.29\%$$

我们就可用概率度 $t=\Delta_p/\mu_p=3\%/2.29\%=1.31$ 来表示极限误差范围的大小，这时要求该市居民户月均用水量在 10 吨以上的比重在 $30\%\pm1.31\mu_p$ 之间。

4.3 抽样估计的方法

抽样估计就是指利用实际调查计算的样本指标值来估计相应的总体指标的数值，即对总体平均数 $\overline{X}$ 、总体成数 P 的推断估计。由于总体指标是表明总体数量特征的参数，所以也称为参数估计。总体参数估计有点估计和区间估计两种。以下分别加以介绍。

4.3.1 总体指标的点估计

点估计的基本特点是，根据总体指标的结构形式设计样本指标作为总体指标的估计量，并以样本指标的实际值直接作为相应总体指标的估计值。例如，以样本平均数的实际值作为相应总体平均数的估计值，以样本成数的实际值作为相应总体成数的估计值等。我们之所以这样做的考虑是基于我们对所研究的总体指标的具体数值虽然不知道，但对它的指标结构形式是清楚的。例如，我们要研究某市的粮食亩产水平，虽然实际的平均亩产的数值是未知的，但平均亩产指标是由总体各单位变量值代数和除以单位数求得的，这个指标的



结构形式则是已知的。很自然可以认为如果抽样调查所取得的样本数据有足够的代表性，那么根据已知的指标结构形式计算样本指标值，便可以作为相应总体指标的估计值。

设 $\bar{x}$ 表示总体平均数 $\overline{X}$ 的估计量，p 表示总体成数 P 的估计量，则有

$$\bar{x}=\overline{X}\ ,\quad p=P$$

式中，$\bar{x}=\dfrac{\sum x}{n}$ 和 $\overline{X}=\dfrac{\sum X}{N}$ 有相同的结构形式，$p=\dfrac{n_1}{n}$ 和 $P=\dfrac{N_1}{N}$ 也有相同的结构形式。再经过实际调查取得样本平均数 $\bar{x}$ 和样本成数 p 的实际值，便可以作为总体平均数 $\overline{X}$ 和总体成数 P 的估计值。例如，我们以样本平均用水量 5 吨作为全市月均用水量的估计值，以样本月用水量 10 吨以上的比重 30%作为全市的估计值等。

对总体指标作估计的时候，总是希望估计是合理的或优良的。那么什么是优良估计的标准呢?

优良估计是从总体上来评价的。其标准有三个方面。

1．无偏性

即以抽样指标估计总体指标要求抽样指标值的平均数等于被估计的总体指标值本身。就是说，虽然每一次的抽样指标值和总体指标值之间都可能有误差，但在多次反复的估计中，各个抽样指标值的平均数应该等于所估计的总体指标值本身，即抽样指标的估计，平均说来是没有偏误的。

抽样平均数的平均数等于总体平均数，抽样成数的平均数等于总体成数，则

$$E(\bar{x})=\overline{X}\ ,\quad E(p)=P$$

这说明以抽样平均数作为总体平均数的估计量，以抽样成数作为总体成数的估计量，是符合无偏性原则的。

2．一致性

以抽样指标估计总体指标要求当样本的单位数充分大时，抽样指标也充分地靠近总体指标。就是说，随着样本单位数 n 的无限增加，抽样指标和未知的总体指标之差的绝对值小于任意小的数，它的可能性也趋近于必然性，即实际上是肯定的。

我们知道，抽样平均数和抽样成数的抽样平均误差与样本单位数的平方根成反比例变化，样本单位数越多则平均误差便越小，当样本单位数接近于总体单位数时，平均误差也就接近于零。也就是说，抽样平均数和抽样成数作为总体平均数和总体成数的估计量是符合一致性原则的。

3．有效性

以抽样指标估计总体指标要求作为优良估计量的方差应该比其他估计量的方差小。例

如，用抽样平均数或总体某一变量值来估计总体平均数，虽然两者都是无偏的，而且在每一次估计中，这两种估计量和总体平均数都可能有离差，但样本平均数更靠近于总体平均数的周围，平均说来，其离差比较小。所以对比说来，抽样平均数是更为有效的估计量。

总体指标点估计方法的优点是简便易行，原理直观，常为实际工作所采用。但也有不足之处，即这种估计没有表明抽样估计的误差，更没有指出误差在一定范围内的概率保证程度有多大。要解决这个问题，必须采用总体指标的区间估计方法。

4.3.2 总体指标的区间估计

区间估计就是以一定的概率保证估计包含总体指标的一个值域，即根据样本指标和抽样平均误差推断总体指标的可能范围。它包括两部分内容：一是这一可能范围的大小；二是总体指标被包含在这个可能范围内的概率。区间估计既说明估计结果的精度，又同时表明这个估计结果的可靠程度、保证程度，所以区间估计是比较科学的，它是本节阐述的重点。

1．抽样估计的置信度

我们已经学习了确定允许的抽样误差范围，从主观愿望说，当然希望抽样调查的结果，即样本指标的估计值都能够被包含在允许的误差范围内，但这并非是能实现的事情。由于抽样指标值随着样本的变动而变动，它本身是个随机变量，因此抽样指标和总体指标的误差仍然是个随机变量，并不能保证误差不超过一定范围的这件事是必然的，而只能给以一定程度的概率保证。抽样估计置信程度就是表明抽样指标和总体指标的误差不超过一定范围的概率保证程度。

所谓概率，就是指在随机事件进行大量试验中，某种事件出现的可能性大小，它通常可以用某种事件出现的频率来表示。抽样估计要求的保证程度就是指抽样误差不超过一定范围的概率大小。可以用以下形式表示：

$$P(|\bar{x}-\overline{X}| \leqslant \Delta_{\bar{x}}) = P_1 + P_2 + \cdots + P_K$$

等式左边括号内的 $|\bar{x}-\overline{X}| \leqslant \Delta_{\bar{x}}$ 表示样本平均数与总体平均数的误差不超过 $\Delta_{\bar{x}}$，$P(|\bar{x}-\overline{X}| \leqslant \Delta_{\bar{x}})$ 则表示误差不超过这一范围的概率。等式右边表示属于这一区间范围内各种样本平均值出现的概率之和。

【例 4-4】

现在仍用例 4-1 来说明抽样估计置信度的概念。已知 4 个工人的日产量分别为 70，90，130，150 件，总平均日产量为 110 件，总体标准差为 22.36 件。用重复抽样的方法从中取 2 人为样本，计算样本平均日产量，并加以整理。平均日产量的分布如表 4-3 所示。

表 4-3　平均日产量的分布

样本平均数 x	70	80	90	100	110	120	130	140	150
频数 f	1	2	1	2	4	2	1	2	1
频率（概率）	1/16	2/16	1/16	2/16	4/16	2/16	1/16	2/16	1/16

我们可以根据以上分布写出平均日产量落在各种区间范围内的概率 P，例如：

$$P(100 \leqslant \bar{x} \leqslant 120) = \frac{2}{16} + \frac{4}{16} + \frac{2}{16} = \frac{1}{2} = 50\%$$

$$P(80 \leqslant \bar{x} \leqslant 140) = \frac{2}{16} + \frac{1}{16} + \frac{2}{16} + \frac{4}{16} + \frac{2}{16} + \frac{1}{16} + \frac{2}{16} = \frac{14}{16} = \frac{7}{8}$$

将上述概率形式变换成抽样误差的形式，即求得抽样平均数与总体平均数误差绝对值不超过一定范围的概率。例如：

$$P(|\bar{x} - \bar{X}| \leqslant 10) = \frac{1}{2}$$

$$P(|\bar{x} - \bar{X}| \leqslant 30) = \frac{7}{8}$$

…

这说明在重复抽样中，抽样平均日产量与总体平均日产量绝对误差不超过 10 件的概率为 1/2，即有 50%的概率保证在一次抽样中使上述误差得以实现。同理，抽样误差不超过 30 件的概率为 7/8 等。由此可见，抽样误差范围和概率是密不可分的，而且抽样误差范围越小，则估计的概率越小；抽样误差范围越大，则估计的概率也越大。

但当总体很大时，要依靠列表来求抽样误差的置信概率几乎是难以做到的。在单位数为 N 的总体中用不重复抽样的方法抽取 n 个单位构成一个样本，则全部的可能样本共有 C_N^n 个，如果包括重复的样本就更多，选到 N^n 个。把全部可能样本平均数编制成如表 4-3 所示的变量数列，然后再将它们绘成图形，即可得到一个钟形图的平滑曲线，如图 4-1 所示。

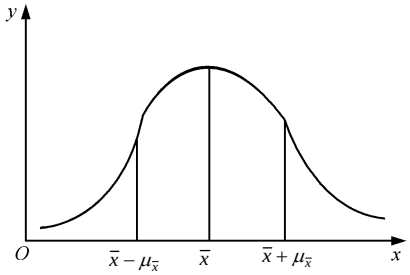


图 4-1　正态分布曲线图

图中的曲线叫正态分布曲线，它表明全部可能样本指标的分布状况。数理统计已证明，在样本单位数足够多（n>30）的条件下，无论总体单位标志值如何分布，其抽样平均数的分布接近于正态分布。从正态分布图中，可以看出这一分布的特点是：一是样本指标是以总体指标为中心的，两边完全对称分布，即样本指标高于与低于总体指标的可能性是完全相等的；二是样本指标越接近于总体指标，其出现的可能性越大，概率就越大。反之，抽样指标越远离总体指标，其出现的可能性越小，概率越小，最终趋于 0。

根据数学证明，在 $\bar{x}-\mu_{\bar{x}}$ 到 $\bar{x}+\mu_{\bar{x}}$（简写为 $\bar{x}\pm\mu_{\bar{x}}$）区间内，这一部分曲线下的面积占曲线下全部面积的 68.27%；在 $\bar{x}-2\mu_{\bar{x}}$ 到 $\bar{x}+2\mu_{\bar{x}}$（简写为 $\bar{x}\pm2\mu_{\bar{x}}$）区间内，这一部分曲线下的面积占曲线下全部面积的 95.45%；在 $\bar{x}-3\mu_{\bar{x}}$ 到 $\bar{x}\pm3\mu_{\bar{x}}$（简写为 $\bar{x}+3\mu_{\bar{x}}$）区间内，这一部分曲线下的面积占曲线下全部面积的 99.73%，如图 4-2 所示。

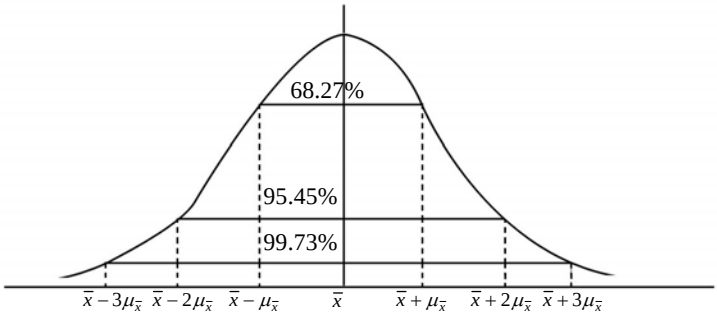


图 4-2　样本指标置信图

若全部可能的样本是 10 000 个，则在 $\bar{x}\pm\mu_{\bar{x}}$ 区间内有 6 827 个样本，在 $\bar{x}\pm2\mu_{\bar{x}}$ 区间内有 9 545 个样本，在 $\bar{x}\pm3\mu_{\bar{x}}$ 区间内有 9 973 个样本。这里的 68.27%，95.45%和 99.73%叫样本在 $\bar{x}\pm\mu_{\bar{x}}$，$\bar{x}\pm2\mu_{\bar{x}}$ 和 $\bar{x}\pm3\mu_{\bar{x}}$ 区间内出现的概率，也叫样本指标在该区间内的可靠程度、保证程度或置信程度。抽样估计的置信程度就是表明样本指标和总体指标的误差不超过一定范围的概率有多大。

由于概率度 $t=|\bar{x}-\bar{X}|/\mu_{\bar{x}}$，所以抽样估计的概率保证程度就是概率度 t 的函数，可表示为 $F(t)$。

当　　t=1 时，$F(t)$=68.27%

　　t=2 时，$F(t)$=95.45%

　　t=3 时，$F(t)$=99.73%

t 越大，误差范围就越大，样本出现的概率或保证程度也就越大；反之则小。根据数学计算，它的不同取值的误差范围，均有其对应的概率或保证程度。将这种函数关系编制成正态分布概率表，给定 t 值，便可以直接从表上查找抽样误差的概率，即估计置信度（见

附录的附表 A-1)。

2．总体指标的区间估计方法

我们理解了估计值、抽样误差范围，以及抽样误差范围的概率保证程度之后，就可以来研究总体指标的区间估计了。总体指标区间估计的基本特点是根据给定的概率保证程度的要求，利用实际抽样资料，指出总体被估计值可能存在的区间范围，而不是直接给出总体指标的估计值。换句话说，对于总体的被估计指标 $\overline{X}$ ，找出样本的两个估计量 x_1 和 x_2，使被估计指标被包含在区间（x_1, x_2）内的概率是 P，即 $P(x_1 \leqslant \overline{X} \leqslant x_2) = 1-\alpha$ ，α 是给定的。我们称区间（x_1, x_2）为总体指标 $\overline{X}$ 的置信区间，其估计置信度为 $1-\alpha$ ，称 α 为显著性水平，x_1 是置信下限，x_2 是置信上限。

如对例 4-3 的居民户均用水量，也可以作如下的区间估计，由于 $t=2$，查表得 $F(t)=0.9545$，即以 95.45%的概率保证，全部居民户均用水量 $\overline{X}$ 在 4.92 ~ 5.08 吨之间。估计置信度 $1-\alpha = 95.45\%$，显著性水平 α = 1–95.45% = 4.55%，它表示全部居民户均用水量包含在 4.92 ~ 5.08 吨区间内有 95.45%的概率，而不被包含在这个区间内的概率有 4.55%，因此做上述区间估计就必须冒不超过 4.55%概率的失败风险。显著性水平正是判别估计可信不可信的一个标准。当你不愿意冒这样大的风险时，可以缩小显著性水平，则置信区间就要扩大，估计的准确性便要降低了。就如例 4-3，当以 99.73%的概率估计全部居民户均用水量时，$t=3$，用水量区间扩大为（4.88，5.12），显著性水平为 1– 99.73% = 0.27%，这就只要冒 0.27%概率的失败危险。

由此可见，总体指标的区间估计必须同时具备估计值、抽样误差范围和概率保证程度三个要素，抽样误差范围决定估计的准确性，而概率保证程度则决定估计的可靠性。在抽样估计的时候自然希望估计的准确性要尽量高些，而估计的可靠性也要尽量大些。但是这两个愿望是矛盾的，对于一个样本，提高了估计准确性的要求，就降低了估计的可靠性；同样，提高了估计可靠性的要求，也就降低了估计的准确性。因此在抽样估计的时候，先对其中的一个要素提出要求，再推算另一个要素的变动情况。例如对估计的准确性提出要求，即要求误差范围不超过给定的标准，来推算估计的可靠性，即概率保证程度。或对估计的可靠性提出要求，即要求给定概率保证程度，来推算可能的误差范围。如果所推算的另一要素（不论是准确性或可靠性）不能满足实际工作的需要就应该采取其他措施，如增加样本单位、改善抽样组织、重新进行抽样，直到符合要求为止。

所以总体指标的区间估计根据所给定的条件不同，而有两种估计方法。一种是根据已经给定的抽样误差范围，求概率保证程度。具体步骤是：首先，抽取样本，计算抽样指标，作为相应总体指标的估计值，并计算样本标准差以推算抽样平均误差。其次，根据给定的抽样极限误差范围，估计总体指标的下限和上限。最后，将抽样误差除以抽样平均误差求出概率度 t 值，再根据 t 值查正态分布概率表求出相应的置信度 $F(t)$，并对总体参数作区间估计。

➪【例 4-5】

从某县农民家庭中随机抽样 100 户调查其年收入情况。农民家庭按年人均纯收入额分组的资料如表 4-4 所示。

表 4-4　某县百户农民家庭户按人均年收入分组资料

按人均收入分组（元）	农户数
9 000 以下	3
9 000～10 000	7
10 000～11 000	28
11 000～12 000	32
12 000～13 000	20
13 000 以上	10
合　计	100

要求全县农民家庭户的人均收入的允许误差范围为 $\Delta_{\bar{x}}=250$ 元，人均收入在 10 000 元以上户数所占比重的估计的误差范围不超过 5%。

- 计算抽样平均数和抽样平均误差，如表 4-5 所示。

表 4-5　某县百户农民家庭户按人均年收入分组抽样指标计算表

按人均收入分组（元）	农户数 f	组中值 x	xf	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
9 000 以下	3	8 500	25 500	8 352 100	25 056 300
9 000～10 000	7	9 500	66 500	3 572 100	25 004 700
10 000～11 000	28	10 500	294 000	792 100	221 788 800
11 000～12 000	32	11 500	36 800	12 100	387 200
12 000～13 000	20	12 500	25 000	1 232 100	24 642 000
13 000 以上	10	13 500	135 000	4 452 100	44 521 000
合　计	100	—	1 139 000	—	141 790 000

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{1139\,000}{100}=11\,390(\text{元}),\quad p=\frac{90}{100}=90\%$$

$$s^2=\frac{\sum(x-\bar{x})^2 f}{\sum f}=\frac{141\,790\,000}{100}=1\,417\,900$$

$$m_{\bar{x}}=\sqrt{\frac{s^2}{n}}=\sqrt{\frac{1\,417\,900}{100}}=119.08,\quad m_p=\sqrt{\frac{p(1-p)}{n}}=\sqrt{\frac{90\%?\ (1\quad 90\%)}{100}}=0.03$$

- 根据给定 $\Delta_{\bar{x}}$ = 250 元，Δ_p = 5%，求总体平均数、总体成数的上下限。

总体平均数下限 $\bar{x}-\Delta_{\bar{x}}$=11 390–250=11 140，上限 $\bar{x}+\Delta_{\bar{x}}$=11 390+250=11 640

总体成数下限 $p-\Delta_p$ = 90%–5%=85%，上限 $p+\Delta_p$ =90%+5%=95%

- 根据 t 值，查概率表得出置信度。

$$t=\frac{\Delta_{\bar{x}}}{\mu_{\bar{x}}}=\frac{250}{119.07}=2.1 \quad \text{查表得} \quad F(t)=0.964\,3$$

$$t=\frac{\Delta_p}{\mu_p}=\frac{5\%}{3\%}=1.67 \quad \text{查表得} \quad F(t)=0.905$$

根据计算，我们得出如下估计：可以 96.43%的概率保证程度估计该县农民家庭户人均年收入在 11 140～11 640 元之间；可以 90.51%的概率保证程度估计该县农民家庭中年人均收入在 10 000 元以上的户数所占比重在 85%～95%之间。

另一种方法是根据给定的置信度要求，来推算抽样极限误差的可能范围。具体步骤是：首先，抽取样本，计算抽样指标，如计算抽样平均数或抽样成数作为总体指标的估计值，并计算样本标准差以推算抽样平均误差。其次，根据给定的置信度 $F(t)$要求，查概率表求得概率度 t 值。最后，根据概率度 t 和抽样平均误差来推算抽样极限误差的可能范围，再根据抽样极限误差求出被估计总体指标的上下限，对总体参数作区间估计。

⇨【例 4-6】

某灯泡厂从一批 10 000 只灯泡中随机抽取 500 只，检验按规定灯泡耐用时数在 850 小时以上者为合格品。经检验，其平均耐用时数为 925 小时，其中合格品 463 只，标准差为 54.3 小时。试以 95.45%的置信度对这批灯泡的平均耐用时数和合格品率进行区间估计（按不重复抽样方法）。

- 根据抽样资料，已知 $\bar{x}$ =925 小时，s=54.3 小时，$p=\frac{463}{500}=92.6\%$。

$$m_{\bar{x}}=\sqrt{\frac{s^2}{n}(1-\frac{n}{N})}=\sqrt{\frac{54.32^2}{500}?\ (1\quad\frac{500}{10\,000})}=2.367$$

$$\mu_p=\sqrt{\frac{p(1-p)}{n}(1-\frac{n}{N})}=\sqrt{\frac{0.926\times(1-0.926)}{500}\times(1-\frac{500}{10\,000})}=0.011\,4$$

- 根据给定的概率置信度 $F(t)$ =0.954 5，查概率分布表得 t =2。
- 计算 $\Delta_{\bar{x}}$，Δ_p，可得出估计区间的上下限。

$$\Delta_{\bar{x}} = t\,\mu_{\bar{x}} = 2 \times 2.367 = 4.734$$

$$\Delta_p = t\,\mu_p = 2 \times 1.14\% = 2.28\%$$

即：　灯泡耐用时数的下限　$\bar{x} - \Delta_{\bar{x}} = 925 - 4.734 \approx 920$（小时）

灯泡耐用时数的上限　$\bar{x} + \Delta_{\bar{x}} = 925 + 4.734 \approx 930$（小时）

灯泡合格率的下限　$p - \Delta_p = 92.6\% - 2.28\% \approx 90.3\%$

灯泡合格率的上限　$p + \Delta_p = 92.6\% + 2.28\% \approx 94.9\%$

上述计算表明，该批灯泡的平均耐用时数在 920～930 小时之间，合格率在 90.3%～94.9%之间，其估计可靠程度为 95.45%。

4.4 抽样组织设计

4.4.1 抽样组织设计的基本原则

抽样推断是根据事先规定的要求而设计的抽样调查组织，并以所获得的这一部分实际资料为基础，进行推理演算做出结论。因此如何科学地设计抽样调查组织，保证随机条件的实现，并且取得最佳的抽样效果，就是一个至关重要的问题。

在抽样设计中，首先要保证随机原则的实现。随机取样是抽样推断的前提，失去这个前提，推断的理论和方法也就失去了存在的意义。从理论上说，随机原则就是要保证总体每一单位都有同等的中选机会，或样本的抽选概率是已知的。但在实践上，如何保证这个原则的实现，需要考虑许多问题。一是要有合适的抽样框。抽样框固然要具备可实施的条件，可以从中抽取样本单位，但仅仅这样是很不够的。必须考虑一个合适的抽样框是否能覆盖总体的所有单位。例如，某城市进行民意调查，如何以该市的电话号码簿名单为抽样框显然是不合适的，因为并不是所有居民户都安装电话，而且安装电话居民户又多数是经济条件较好的上层人物，从这里取得的样本资料是很难说具有全市的代表性。还要考虑抽样框中抽样单位与总体单位的对应问题。在实践中发生不一致的问题也不是少见的。有的是多个抽样单位对应一个总体单位。例如，调查学校学生家庭情况，以学生名单为抽样框，在学生名单中可能有两个或更多的学生属于同一个家庭。也有的是一个抽样单位对应几个总体单位。例如，人口调查中以住户列表为抽样框，每一住户就包括许多人口。像这类抽样很可能造成总体单位中选机会不均等，应该注意加以调整。二是取样的实施问题。当总体单位数很大甚至无限的情况下，要保证总体每单位中选的机会均等绝非是简单的工作。在设计中要考虑将总体各单位加以分类、排队或分阶段等措施，尽量保证随机原则的实现。

其次，要考虑样本容量和结构问题。样本的容量究竟要多大才算是合适的？例如，在粮食产量调查中，要调查多少亩才能反映全省几千万亩播种面积的亩产水平？在民意测验



中，要调查多少人才能反映全国十几亿人口的意见等。调查单位多了会增加组织抽样的负担，甚至造成不必要的浪费；但调查单位太少又不能够有效地反映情况，直接影响着推断的效果。样本的容量取决于对抽样结果的用途而不同，很难给出一个绝对的标准。但在抽样设计时应该重视研究现象的差异、误差的要求和样本容量之间的关系，做出适当的选择。对相同的样本容量，还有容量的结构问题，例如一个县要求抽取 500 亩播种面积，可以是先抽 5 个村，然后每村抽 100 亩，也可以是先抽 10 个村，然后每村抽 50 亩等。样本容量的结构不同，所产生的效果也不同。抽样设计应该善于评价而且有效利用由于调整样本结构而产生的效果。

再次，关于抽样的组织形式问题。要认识到不同的抽样组织形式，会有不同的抽样误差，因而就有不同的效果。一种科学的组织形式往往有可能以更少的样本单位数，取得更好的抽样效果。在抽样设计时必须充分利用已经掌握的辅助信息，对总体单位加以预处理，并采取合适的组织形式取样。例如，粮食生产按地理条件分类，并分类取样；或按历史单产资料、当年估产资料，将各单位顺序排队，并等距取样等，都能收到更好的抽样效果。还应该指出，即使是同一种抽样组织形式，由于采用的分类标志不同、群体的划分不同等，仍然会产生不同的效果。因此应该认真细致地估计不同组织形式和不同抽样方法的抽样误差，并进行对比分析，从中选择有效的切实可行的抽样方案。

在抽样设计中还必须重视调查费用这个基本因素。调查费用指收集数据费、数据处理费、制表费、工作机关管理费、出版费等。实际上任何一项抽样调查都是在一定费用的限制条件下进行的，抽样设计应该力求费用节省的方案。在设计方案中我们还要注意到，提高精确度的要求和节省费用的要求并非一致，有时是相互矛盾的。抽样误差要求越小，则调查费用往往越大，因此并非抽样误差越小的方案便是越好的方案，许多情况是允许一定范围的误差就能够满足分析研究的要求。我们的任务是在一定误差的要求下选择费用最小的方案；或在一定的费用开支条件下，选择误差最小的方案。

本节主要介绍必要抽样数目的确定和几种常用的抽样组织形式。

4.4.2　必要抽样数目的确定

抽样数目是决定抽样误差大小的直接因素，因此，在组织抽样调查时，必须事先确定抽样单位数目。确定必要抽样数目的原则是在保证预期的抽样推断可靠程度的要求下，尽量减少样本单位数目。这是因为，虽然抽取的单位数越多，样本的代表性越大，抽样误差越小，抽样推断就越可靠，但是抽取的单位数过多，会增加不必要的人力、物力和费用开支，造成浪费，而且还会影响资料提供的及时性。所以，应当在抽样调查之前，根据调查对象的特点和研究目的要求，做出科学的设计，确定出必要的抽样数目，使其既不浪费人力、物力、财力，又能取得较好的抽样推断效果。

1．影响必要抽样数目的因素

（1）总体被研究标志的变异程度

总体的标志变异程度大，需多抽样本单位；反之则可少抽。具体来说，可以看总体各单位被研究标志方差 σ^2 或 P（1–P）的大小。方差大，抽样数目应确定得多些；相反则少些。

（2）对推断精确度的要求

对推断精确度的要求越高，允许误差 Δ 越小，这时就需要多抽取一些样本单位；如果对推断精确度的要求较低，允许误差 Δ 大一些，则可以少抽取一些样本单位。

（3）对推断可靠性的要求

抽样推断的可靠程度也就是概率，概率与 t 值有关。如果要求可靠程度高，t 值就大，需要多抽取样本单位；反之，如果要求的可靠程度低，概率小，t 值小，则可以少抽取样本单位。

（4）抽样调查的组织方式和方法

同一对象要求有同样的精确度和保证程度，用机械抽样和类型抽样，抽样数目可定得少些。若用纯随机和整群抽样方式，抽样数目就要定得多些。至于用重复抽样或不重复抽样方法，后者的抽样数目可确定得少些。

（5）人力、物力和财力的允许条件

从以上因素考虑的抽样数目，还应结合调查的人力、物力和财力的具体情况做适当调整，然后再最后确定。

在抽样调查前，调查者通常要根据调查对象的特点和研究目的，提出两条主要要求：

- 抽样调查的误差范围或允许误差不得大于多少，这是规定了误差范围 Δ 的值；
- 抽样推断的结果要有多大的保证程度，这是规定了概率度 t 值。

可见，必要抽样数目的计算公式，是从 $\Delta = t\mu$ 这个公式推导出来的。

2．简单随机抽样必要抽样数目的计算公式

（1）在重复抽样条件下

推断总体平均数所需要的抽样数目：

$$n_{\bar{x}} = \frac{t^2\sigma^2}{\Delta_{\bar{x}}^2} \tag{4-11}$$

推断总体成数所需要的抽样数目：

$$n_p = \frac{t^2P(1-P)}{\Delta_P^2} \tag{4-12}$$

（2）在不重复抽样条件下

推断总体平均数所需要的抽样数目：

$$n_{\bar{x}}=\frac{t^2\sigma^2N}{N\Delta_{\bar{x}}^2+t^2\sigma^2} \tag{4-13}$$

推断总体成数所需要的抽样数目：

$$n_P=\frac{t^2P(1-P)N}{N\Delta_P^2+t^2P(1-P)} \tag{4-14}$$

需要说明的是：

- 在实际工作中，由于抽样比例一般很小（即 n/N），虽然采用的是不重复抽样，但仍按重复抽样的公式来计算必要的抽样数目；
- 公式中的σ和 P 一般都未知，也没有样本数据可代替，通常是利用过去同类调查的数据计算，或测试以取得所需数据；
- 根据平均数的公式和成数的公式所计算出的必要抽样数目往往不相等，有时甚至相差很大，为了保证抽样推断的准确程度，则应选用其中较大的样本数值。

【例 4-7】

假定某乡有农户 18 000 户，在某次调查中采用重复的纯随机方式进行抽样，要求人均收入的极限误差控制在 150 元内，把握程度为 95.45%，该抽取多少农户？如果极限抽样误差要求控制在 75 元内，应抽取多少户？（注：根据以往资料，全乡人均收入标准差为 1 500 元。）

（1）采用重复抽样公式计算（查表得 t=2）

当极限误差 $\Delta_{\bar{x}}\leqslant 150$ 元时，有：

$$n_{\bar{x}}=\frac{t^2\sigma^2}{\Delta_{\bar{x}}^2}=\frac{2^2\times 1\ 500^2}{150^2}=400(户)$$

当极限误差 $\Delta_{\bar{x}}\leqslant 75$ 元时，有：

$$n_{\bar{x}}=\frac{t^2\sigma^2}{\Delta_{\bar{x}}^2}=\frac{2^2\times 1\ 500^2}{75^2}=1\ 600(户)$$

可见，在重复抽样中，极限误差缩小一半（即为原来的 1/2）时，必须把样本容量增到 4 倍。

（2）采用不重复抽样公式计算

当极限误差 $\Delta_{\bar{x}}\leqslant 150$ 元时，有：

$$n_{\bar{x}}=\frac{t^2\sigma^2N}{N\Delta_{\bar{x}}^2+t^2\sigma^2}=\frac{2^2\times1\ 500^2\times18\ 000}{18\ 000\times150^2+2^2\times1\ 500^2}\approx392(\text{户})$$

当极限抽样误差缩小一半（即为原来的 1/2）时，有：

$$n_{\bar{x}}=\frac{t^2\sigma^2N}{N\Delta_{\bar{x}}^2+t^2\sigma^2}=\frac{2^2\times1\ 500^2\times18\ 000}{18\ 000\times75^2+2^2\times1\ 500^2}\approx1\ 470(\text{户})$$

如果是采用其他抽样组织方式，则公式略有不同。可参见下面介绍的几种常用的抽样组织形式的抽样平均误差的计算方法。

4.4.3 常用的抽样组织形式

1．简单随机抽样

简单随机抽样也叫纯随机抽样，它对总体单位不作任何分类排队，而是直接从总体中随机抽取一部分单位来组成样本的抽样组织方式。抽样的原理是以纯随机抽样为基础阐述的，它是最简单最基本的抽样组织方式，是设计其他复杂抽样组织方式的基础。简单随机抽样适用于总体各部分都具有相同的分布且单位数较少的情况。

采用简单随机抽样方式抽取样本，先要将总体各个单位进行编码，然后按随机原则抽取若干数码，所有中选数码所对应的单位即构成样本。具体做法如下。

（1）抽签法

当给总体各单位编号后，把号码写在结构均匀的签上，将签均匀后即可以从中抽取。采用这种方法简便易行，然而对较大的总体来说，编号作签工作量很大，混匀有困难，因此，这种方法的应用有一定局限性。

（2）随机数字法

随机数字可以借助于计算机获得，也可应用随机数表（见附录的附表 A-2）。表中数字是按照完全随机的方法产生的，从 0 至 9 中任何一个（一组）数字出现的概率与其他（组）数字出现的概率完全相同。随机数表的使用要遵守随机原则。首先，给总体编号，据编号的最大位数确定将要使用随机数表的列数，然后从表中任意一列、任意一行开始，由纵向或横向画线取数，遇到属于总体单位编号范围内的数组就确定为样本单位，然后继续往下找。如果要求不重复抽样时，遇到重复出现的数字（组）就弃之，直到取足要求的单位数为止。

例如，有一个 80 个单位构成的总体，现在要从中抽取 8 个单位。首先将总体各单位从 1 至 80 编号，故最高位是两位数，然后随机确定行、列开始取数，假定从第 3 行第 6 列的数字 06 开始取样，沿行抽取，于是有 06，72，17，10，25，21，31，75 这 8 组数字所对应的单位即构成所需的样本。



2．类型抽样

类型抽样又称分层抽样。它的特点是先对总体各单位按主要标志加以分组，然后再从各组中按随机的原则抽选一定单位构成样本。

设总体由 N 个单位构成，把总体划分为 K 组，使 $N=N_1+N_2+\cdots+N_K$，然后从每组的 N_i 单位中抽取 n_i 单位构成样本容量为 n 的样本，使 $n=n_1+n_2+\cdots+n_K$，这种抽样方法称为类型抽样。

通过分类，可以把总体中标志值比较接近的单位归为一组，减少各组内的差异程度，再从各组抽取样本单位就有更大的代表性，因而抽样误差也就相对缩小了。在总体单位标志值大小悬殊的情况下，运用类型抽样比简单随机抽样可以得到比较准确的效果。这在实际工作中得到了广泛的应用，如农产量抽样按地理条件分组，职工家计调查按国民经济部门分组，产品质量抽检按加工车床型号分组等，都收到明显的效果。

类型抽样的样本单位数在各类型之间的分配有三种方法：

（1）等数分配类型抽样法

它是在各类型组中分配同等的样本单位数的方法。这种方法只在各类型的总体单位数相等或差异不大时才使用，运用这种分配方法可使综合计算比较简单。

（2）等比例类型抽样法

它是按照类型大小以相等的比例分配样本的方法。由于是按有关的主要标志分类（组），各组的单位数一般不同。类型抽样通常是按各组总体单位数占总体单位数的一定比例来抽取样本的，单位数较多的组应该多取样，单位数较少的组则少取样，保持各组样本单位数与样本总容量之比等于各组总体单位数与总体单位数之比。

即：

$$\frac{n_1}{N_1}=\frac{n_2}{N_2}=\cdots=\frac{n_K}{N_K}=\frac{n}{N}$$

所以各组的样本单位数应为：

$$n_i=\frac{nN_i}{N}(i=1,2,\cdots,K)$$

采用等比例抽样是为了使样本的结构接近总体的结构，避免样本平均数由于各组比重差异而引起的误差。由于等比例类型抽样法对样本单位的分配比较合理，在实际工作中应用较多。

（3）不等比例类型抽样法

它是在各类型组中按不同的比例分配样本单位的方法，也叫最优分配法。当各类型组的单位数相差悬殊或标志变异程度相差较大时，上述两种样本分配方法的抽样效果就差些。此时宜采用不等比例类型抽样法抽取样本，即标志变异程度大或单位数多的组的抽样比例

可高一些，多抽一些单位；标志变异程度小或单位数少的组的抽样比例可低一些，少抽一些单位。

3．等距抽样

等距抽样也称机械抽样或系统抽样。它先按某一标志对总体各单位进行排队，然后依一定顺序和间隔来抽取样本单位的一种抽样组织。这种方法可以保证所取得的样本单位比较均匀地分布在总体的各个部分，有较高的代表性。

作为总体各单位顺序排列的标志，可以是无关标志也可以是有关标志。所谓无关标志，是指和单位标志值的大小无关或不起主要的影响作用。例如，工业产品质量抽查按时间顺序取样，农产量抽样调查按田间的地理顺序取样，居民家计调查按街道的门牌号码抽取调查户等。

在对总体各单位的变异情况有所了解的情况下，也可以采用有关标志进行总体单位排队。所谓有关标志，是指作为排队顺序的标志和单位标志值的大小有密切的关系。例如农产量抽样调查，利用各县或乡近三年亩产量或当年估计亩产量排队，抽取调查单位。又如职工家计调查，按上年职工平均工资排队，抽取调查企业或调查户等。按有关标志排队实质上是运用类型抽样的一些特点，有利于提高样本的代表性。

在等距抽样中，不论是无关标志还是有关标志排队，都要注意避免样本间隔与现象本身的周期性节奏相重合，引起系统性误差的影响。例如农产量抽样调查，样本点的抽样间隔不宜和田间的长度相等；工业产品质量抽查，产品抽样时间间隔不宜和上下班时间一致，以免发生系统性的偏差，影响样本的代表性。

等距抽样的抽样平均误差和标志排列的顺序有关，情况比较复杂。一般地，如果用来排队的标志是无关标志，而且随机起点取样，那么它的抽样误差就十分接近简单随机抽样的误差。为了简便起见，可以采用简单随机抽样误差公式来近似反映。

4．整群抽样

整群抽样也称集团抽样。它是将总体各单位划分成许多群，然后从中随机抽取部分群，对中选群的所有单位进行全面调查的抽样组织形式。例如，要调查家庭副业发展情况，不是直接抽取居民户，而是以村为单位，从中抽取若干村，然后对中选村的全体居民户进行调查，这样就方便多了。

整群抽样的好处是组织工作方便，确定一群便可以调查许多单位。但是正由于抽样单位比较集中，限制了样本在总体分配的均匀性，所以代表性较低，抽样误差较大。在实际工作中，采用整群抽样方法通常都要增加一些样本单位，以减少抽样误差，提高估计准确性。

4.4.4 抽样方案的检查

抽样方案的设计，或由于实际情况有变化，根据历史资料已过时不宜使用；或由于考

虑不周在设计时发生失误、未被发现等原因，不能保证抽样结果都能获得有充分代表性的数据。因此在设计方案实施之前必须经过检查，用试点的结果来验证设计方案的准确和可行，然后才能正式推广使用。方案的检查主要有准确性检查和代表性检查两方面。

1．准确性的检查

所谓准确性的检查，就是以方案所要求的允许误差范围为标准。用已掌握的资料检查其在一定概率保证下，实际的极限误差是否超过方案所允许的误差范围。即要求 $\Delta_{\bar{x}}$ 或 Δ_p 小于或等于所要求的允许误差范围。如果检查结果实际的 $\Delta_{\bar{x}}$ 或 Δ_p 没有超过规定的范围，便认为方案的设计符合准确性的要求，可以付诸实施。反之，如果实际的 $\Delta_{\bar{x}}$ 或 Δ_p 超过了规定的标准，说明方案的准确性不符合要求，就应该对抽样方案进行认真的分析，逐步检查。若不存在技术性的差错，就要增加样本容量，对方案作必要的修正，然后再进行检查，直到符合准确性方面的要求为止。

例如，某县设计的农产量抽样调查方案，样本容量为 15 个村，用等距抽样方法，要求抽样误差范围不超过 10 千克，其置信度为 95%。检查时，先对该县用等距的方法抽 15 个村登记各村的预计产量，并计算其单位面积产量及其抽样平均误差。求得 $\mu = 4.5$ 公斤，实测 $\Delta = t\mu = 1.96 \times 4.5 = 8.82$ 千克。将实测的极限误差 8.82 千克与方案规定的允许误差范围 10 千克作比较，8.82 千克<10 千克，说明方案符合准确性的要求，可以用 15 个村的平均单位面积产量来推断全县的平均单位面积产量水平。

2．代表性的检查

所谓代表性的检查，是指方案中的样本指标与过去已掌握的总体同一指标 $\bar{x}$ 或 P 进行对比，视其比率是否超过规定的要求。按我国制度规定农产量的比率不超过 ±2%，居民收入的比率不超过 ±3%，即：

$$98\% \leqslant \frac{\text{样本平均单产 } \bar{x}}{\text{总体平均单产 } \bar{X}} \leqslant 102\%$$

$$97\% \leqslant \frac{\text{样本平均收入 } \bar{x}}{\text{总体平均收入 } \bar{X}} \leqslant 103\%$$

就认为该方案所具有的代表性令人满意，可以实施；否则，即表示代表性不足，需要对方案进行多方面的检查，做出修正。如果修正后代表性仍不符合要求，就必须增加样本容量以求得满意的代表性。

例如，假定某县所抽的 15 个村计算的平均单位面积产量为 355 公斤，而全县的平均单位面积产量为 350 公斤。那么用 15 个村的平均单位面积产量来代表全县的平均单位面积产量是否具有代表性呢？我们可以计算平均单位面积产量的代表性检查指标：

$$\frac{\text{样本平均单产 } \bar{x}}{\text{总体平均单产 } \overline{X}}=\frac{355}{350}=101.4\%$$

由于 101.42%<102%，没有超过国家制度规定的范围，就可以认为该方案在代表性方面也是符合要求的。

我国农产量和居民家计抽样都是定点调查，在选定调查点之后，连续几年调查。由于总体的情况不断有所变化，因此样本的准确性和代表性连续各年都要进行检查，以保证抽样资料的准确可靠。

本章小结

1. 抽样推断是按照随机原则，从全部研究总体中抽取一部分单位进行调查，并依据所获得的数据对总体的某一数量特征做出具有一定可靠程度的估计与推断的一种统计方法。抽样推断的全过程，就是抽样调查。

2. 与全面调查相比，抽样调查能节省人力、费用和时间，而且比较灵活；在有些情况下，抽样调查的结果比全面调查要准确；抽选部分单位时要遵循随机原则；抽样调查会产生抽样误差，但抽样误差可以计算，并且加以控制。

3. 抽样调查包括：简单随机抽样、分层抽样、等距抽样、整群抽样。各种抽样方式有各自的优缺点，我们可以根据所选样本的特点，选择合适的抽样方法。

4. 抽样误差指样本指标与被它估计未知的总体参数（总体特征值）之差。抽样中误差的来源主要有两方面：一类是登记性误差，另一类是代表性误差。

5. 参数估计可以分为点估计和区间估计。点估计是由样本统计量的一个具体数值估计总体参数。区间估计是样本数值依据一定概率给出总体参数的一个可能位于的区间。

6. 对总体参数进行估计时，应根据已知的条件选择不同的统计量，否则，参数的置信区间的公式是不能成立的。

7. 确定适当的样本容量，是实际抽样之前必须解决的问题。样本容量的多少决定了估计的精确度，同时也影响到抽样的费用。应该根据抽样的目的，选择适当的公式进行计划，做到“心中有数”。

复习思考题

一、概念题

随机原则 样本 简单随机抽样 重复抽样 不重复抽样 类型抽样 整群抽样 点估计 区间估计 抽样平均误差 抽样误差 概率度 抽样极限误差

二、简答题

1. 什么是抽样推断？它有哪些基本的特点？
2. 什么是重复抽样和不重复抽样？不同的抽样方法怎样影响着抽样推断的结果？
3. 什么是抽样误差？为什么它不同于登记误差和系统误差？抽样误差的大小受哪些因素的影响？
4. 怎样理解抽样平均误差就是抽样平均数（或抽样成数）的标准差？它怎样和总体平均数（或总体成数）联系起来，从而可以反映抽样误差的一般水平？
5. 参数估计的优良标准是什么？抽样平均数和抽样成数估计是否符合优良估计标准？试加以说明。
6. 什么是概率度？什么是置信度？这两者有什么关系？
7. 类型抽样中的分组和整群抽样中的分群有什么不同意义和不同要求？
8. 进行简单随机重复抽样，假定抽样单位增加 3 倍，则抽样平均误差将发生如何变化？如果要求抽样误差减少 20%，其样本单位数应如何调整？

三、练习题

1．判断题（把“√”或“×”填在题后的括号里）

（1）随机抽样就是随意抽样。（　　）

（2）某企业在调查本厂的产品质量时，有意把管理较差的某车间的产品不算在内。这种做法必将导致系统性偏差。（　　）

（3）一个全及总体可能抽取很多个样本总体。（　　）

（4）抽样误差产生的原因是抽样调查时违反了随机原则。（　　）

（5）抽样平均误差就是总体指标的标准差。（　　）

（6）极限误差就是最大的抽样误差，因此，总体指标必然落在样本指标和极限误差共同构成的区间之内。（　　）

（7）计算抽样平均误差，当缺少总体方差资料时，可以用样本方差来代替。（　　）

（8）抽样平均误差、总体标准差和样本容量的关系可用公式表达，因此在统计实践中，为了降低抽样平均误差，可缩小总体标准差或增大样本容量来达到。（　　）

（9）重复抽样误差一定大于不重复抽样误差。（　　）

（10）整群抽样为了降低抽样平均误差，在总体分群时注意增大群内方差，缩小群间方差。

（　　）

（11）当全及总体单位数很大时，重复抽样和不重复抽样计算的抽样平均误差相差无几。

（　　）

（12）类型抽样应尽量缩小组间标志值变异，增大组内标志值变异，从而降低影响抽样误差的总方差。（　　）

2．单选题

（1）抽样调查的主要目的在于（　　）。

A. 计算和控制误差　　B. 了解总体单位情况

C. 用样本来推断总体　　D. 对调查单位作深入的研究

（2）抽样调查所必须遵循的基本原则是（　　）。

A. 随意原则　　B. 可比性原则　　C. 准确性原则　　D. 随机原则

（3）下列属于抽样调查的事项有（　　）。

A. 为了测定车间的工时损失，对车间的每三班工人中的第一班工人进行调查

B. 为了解某大学生食堂卫生状况，对该校的一个食堂进行了调查

C. 对某城市居民1%的家庭调查，以便研究该城市居民的消费水平

D. 对某公司三个分厂中的第一个分厂进行调查，以便研究该工厂的能源利用效果

（4）无偏性是指（　　）。

A. 抽样指标等于总体指标　　B. 样本平均数的平均数等于总体平均数

C. 样本平均数等于总体平均数　　D. 样本成数等于总体成数

（5）一致性是指当样本的单位数充分大时，抽样指标（　　）。

A. 小于总体指标　　B. 等于总体指标

C. 大于总体指标　　D. 充分接近总体指标

（6）有效性是指作为优良估计量的方差与其他估计量的方差相比，有（　　）。

A. 前者小于后者　　B. 前者大于后者

C. 两者相等　　D. 两者不等

（7）能够事先加以计算和控制的误差是（　　）。

A. 抽样误差　　B. 登记误差　　C. 代表性误差　　D. 系统性误差

（8）对两个工厂工人平均工资进行不重复的随机抽样调查，抽查的工人人数一样，两工厂工人工资方差相同，但第二个厂工人数比第一个厂工人数整整多一倍。抽样平均误差（　　）。

A. 第一工厂大　B. 第二个工厂大　C. 两工厂一样大　　D. 无法做出结论

（9）抽样平均误差是指抽样平均数（或抽样成数）的（　　）。

A. 平均数　　B. 平均差　　C. 标准差　　D. 标准差系数

（10）在同样情况下，不重复抽样的抽样平均误差与重复抽样的抽样平均误差相比，是（　　）。

A. 两者相等　　B. 两者不等　　C. 前者小于后者　　D. 前者大于后者

（11）反映抽样指标与总体指标之间抽样的可能范围的指标是（　　）。

A. 抽样平均误差　　B. 抽样误差系数

C. 概率度　　D. 抽样极限误差

（12）在下列情况下，计算不重复抽样的抽样平均误差可以采用重复抽样公式（　　）。

A. 总体单位数很多

B. 抽样单位数很少

C. 抽样单位数对总体单位数的比重很小

D. 抽样单位数对总体单位数的比重较大

（13）在进行纯随机重复抽样时，为使抽样平均误差减少 25%，则抽样单位数应（　　）。

A. 增加 25%　　B. 增加 78%　　C. 增加 1.78%　　D. 减少 25%

3．多选题

（1）抽样调查是（　　）。

A. 收集资料的方法　　B. 推断方法　　C. 全面调查方法

D. 典型调查方法　　E. 非全面调查方法

（2）抽样调查的特点是（　　）。

A. 以部分推为全体　　B. 按随机原则抽取单位

C. 抽样调查的目的在于推断有关总体指标

D. 抽样调查的目的在于推断有关总体指标

E. 抽样调查的目的在于了解总体的基本情况

（3）抽样调查可用于（　　）。

A. 有破坏性的调查和推断　　B. 较大规模总体或无限总体的调查和推断

C. 调查效果的提高　　D. 检查和补充全面调查资料

E. 产品的质量检验和控制

（4）从总体中可以抽选一系列样本，所以（　　）。

A. 总体指标是随机变量　　B. 样本指标是随机变量

C. 抽样指标是样本变量的函数　　D. 总体指标是唯一确定的

E. 抽样指标是唯一确定的

（5）抽样误差是（　　）。

A. 抽样估计值与未知的总体真值之差

B. 抽样过程中的偶然因素引起的

C. 抽样过程中的随机因素引起的

D. 指调查中产生的系统性误差

E. 偶然的代表性误差

（6）用抽样指标估计总体指标时，所谓优良的估计应具有（　　）。

A. 无偏性　　B. 一致性　　C. 有效性

D. 准确性　　E. 客观性

（7）抽样推断中的抽样误差是（　　）。

A. 抽样估计值与总体参数值之差　B. 不可避免的

C. 可以事先计算出来的　　D. 可以加以控制的

E. 可以用改进调查方法的办法消除的

（8）影响抽样误差的因素有（　　）。

A. 抽样方法　　B. 样本中各单位标志的差异程度

C. 全及总体各单位标志的差异程度　D. 抽样调查的组织形式

E. 样本容量

（9）抽样平均误差是（　　）。

A. 反映样本指标与总体指标的平均误差程度

B. 样本指标的标准差　　C. 样本指标的平均差

D. 计算抽样极限误差的衡量尺度　E. 样本指标的平均数

（10）在其他条件不变的情况下，抽样极限误差的大小和可靠性的关系是（　　）。

A. 允许误差范围愈小，可靠性愈大

B. 允许误差范围愈小，可靠性愈小

C. 允许误差范围愈大，可靠性愈大

D. 成正比关系

E. 成反比关系

（11）在一定的误差范围要求下（　　）。

A. 概率度大，要求可靠性低，抽样数目相应要多

B. 概率度大，要求可靠性高，抽样数目相应要多

C. 概率度小，要求可靠性低，抽样数目相应要少

D. 概率度小，要求可靠性高，抽样数目相应要少

E. 概率度小，要求可靠性低，抽样数目相应要多

（12）在抽样调查中应用的抽样误差指标有（　　）。

A. 抽样实际误差　　B. 抽样平均误差

C. 抽样误差算术平均数　　D. 抽样极限误差

E. 抽样误差的概率度

（13）影响样本容量大小的因素是（　　）。

A. 抽样的组织形式　　B. 样本的抽取方法

C. 总体标准差大小　　D. 抽样估计的可靠程度

E. 允许误差的大小

（14）计算抽样平均误差时若缺乏全及总体标准差或全及总体成数，可用下述资料代替（　　）。

A. 过去抽样调查所得的有关资料　B. 试验性调查所得的有关资料

C. 重点调查所得的有关资料　D. 样本资料

E. 过去全面调查所得的有关资料

（15）抽样的基本组织形式有（　　）。

A. 纯随机抽样　B. 机械抽样　C. 分层抽样

D. 整群抽样　E. 阶段抽样

4．计算题

（1）某电子产品使用寿命在 3 000 小时以下为不合格品，现在用简单随机抽样方法，从 5 000 个产品中抽取 100 个对其使用寿命进行调查。其结果如表 4-6 所示。

表 4-6　电子产品的使用寿命

使用寿命（小时）	产品个数
3 000 以下	2
3 000～4 000	30
4 000～5 000	50
5 000 以上	18
合　计	100

根据以上资料，要求：

1）按重复抽样和不重复抽样计算该产品平均寿命的抽样平均误差。

2）按重复抽样和不重复抽样计算该产品合格率的抽样平均误差。

3）根据重复抽样计算的抽样平均误差，以 68.27%的概率保证程度对该产品的平均使用寿命和合格率进行区间估计。

（2）某外贸公司出口一种茶叶，规定每包规格不低于 150 克，现在用不重复抽样的方法抽取其中 1%进行检验。其结果如表 4-7 所示。

表 4-7　茶叶检验结果

每包重量（克）	包数
148～149	10
149～150	20
150～151	50
151～152	20

要求：

1）以 99.73%的概率估计这批茶叶平均每包重量的范围，以便确定平均重量是否达到规格要求。

2）以同样的概率保证估计这批茶叶合格率范围。

（3）某地区有 1 000 家商店，按大、中、小型分为三类，其商店数量分别为 N_1=200，N_2=300，N_3=500。今按比例分配抽取一个容量为 n=100 的类型随机样本，平均年营业额（单位：万元）分别为 x_1=120，x_2=75，x_3=40，各类的样本方差分别为 44，18，5。试求该地区平均每家每年营业额的置信度为 95%的置信区间。

（4）某企业有 3 000 名职工，该企业想估计职工们上下班花在路途上的平均时间。以置信度为 99%的置信区间进行估计，并使估计值处在真正平均值附近 1 分钟的误差范围之内，一个先前抽取的小样本给出的标准差为 4.3 分钟。试问应抽取多大的样本?

（5）一个市场分析人员想知道：为了确定某地区看过某种报纸广告的家庭占多大成数，需要从该区抽选多少家庭作样本。这个居民区共有 1000 户，分析人员希望以 95%的置信度对这个成数做出估计，并使估计值处在真正成数附近 0.05 范围之内。在一个先前抽取的样本中，有 25%的家庭看过这种广告。试问应取多大的样本?

（6）一批货物（1 800 箱，24 件/箱）运抵仓库，随机抽取 2%进行检验，获得如表 4-8 所示的资料。试计算抽样平均误差。

表 4-8　货物检验结果

平均每件重量（克）	抽箱数目（箱）
500～540	3
540～580	5
580～620	6
620～660	10
660～700	7
700～740	5
合　计	36

CHAPTER 5

第 5 章　相关分析与回归分析

章首引例

实际生活中，我们可以列举很多客观现象之间存在的相互关系、相互影响、相互制约的案例，例如商品销售价格与利润额、施肥量与农作物的产量、工资增长与劳动生产率、胶卷的出厂时间长短和胶卷的感光速率变化之间的关系。从数量上研究这些现象的相互依存关系，分析现象变动的影响因素和作用程度，在实际生活中很有用。例如，在工业生产中，通过对影响产品成本的各种因素分析，构建一个成本计量模型，以实现成本控制的目的；在农业生产中，通过观察和研究施肥量、密植程度、耕作深度等各因素对农作物产量的影响，建立施肥量、密植程度、耕作深度和农作物产量的回归模型，在一定范围内确定这些变量之间的数学关系，从而为提高农作物产量提供可靠分析。

相关分析与回归分析是处理变量之间关系的一种统计分析方法。对于上述几对变量，我们关心的问题是变量间是否存在关系，关系密切程度如何，它们的具体形式是什么，怎样根据一个变量（或者多个变量）的变动来估计另一个变量的变动，如何用计算机处理和实现这些结果。这是本章要解决的主要问题。

本章学习目标

- ☑ 了解相关分析的意义、作用和种类；
- ☑ 理解相关分析的判断以及相关分析与回归分析的区别和联系；
- ☑ 熟练掌握直线相关系数、一元线性回归模型以及估计标准误差的计算。

5.1 相关分析的意义和内容

5.1.1 相关分析的概念

在自然界和社会现象中，任何现象都不是孤立的，而是普遍联系和相互制约的。现象间的普遍联系、相互制约往往表现为相互依存的关系。这种依存关系通常有两种类型，即函数关系和相关关系。

1．函数关系

函数关系是指现象间存在的一种十分严格的数量依存关系。在这种关系中，某个现象的数值发生变化，都有另一个现象的确定值与它相对应，这种关系可以用数学函数式反映出来。例如，圆的面积随半径的变化而变化。每给定一个圆的半径就有一个唯一确定的圆面积和它对应，面积是半径的函数。在社会经济现象中，同样也存在着这种关系。例如，商品销售额=销售量×销售价格，当销售价格不变时，销售量发生变化，就有一个确定的销售额与它对应，销售额是销售量的函数。

2．相关关系

相关关系是指现象间确实存在的，但是非确定性的相互依存关系。这种关系的特征是：一种现象发生变化，会引起另一种现象的变化，但这种变动关系不是唯一确定的，它可以有多种不同的数量表现。这意味着一个变量虽然受另外一个变量影响，却并不由这一个变量完全确定。例如，粮食平均亩产和施肥量之间存在着一定的依存关系，即随着施肥量的增加，平均亩产一般也会相应地增加，但其增加情况不是唯一确定的。因为平均亩产除受施肥量影响外，还受种子、土壤、气温、雨量、密植程度等因素的影响，这种关系就是相关关系。在许多社会经济现象中都存在着这种相关关系，如提高劳动生产率会使成本降低、利润增加等。

函数关系与相关关系虽然有明显的区别，但两者之间并无严格的界限。由于存在测算误差及研究深度等原因，函数关系在实际中往往通过相关关系表现出来。而在研究相关关系时，为了找出现象间数量关系的内在联系和表现形式，又常需要借助于函数关系的形式加以描述。从这个角度可以说，相关关系是相关分析的研究对象，函数关系是相关分析的工具。相关分析就是要通过对大量数字资料的观察，消除偶然因素的影响，探求现象之间相关关系的密切程度和表现形式。由此我们把研究现象之间相关关系的理论和方法称做相关分析法。

小提示：生活中还有哪些指标（变量）之间的关系是体现为函数关系的？又有哪些指标（变量）之间的关系是体现为相关关系的？

在相关分析中，若相关现象之间存在着一定的因果关系，通常把起决定作用的变量作为自变量，一般用 x 表示，把受自变量影响而相应变化的变量作为因变量，一般用 y 表示。例如，研究劳动生产率与利润之间的关系时，劳动生产率为自变量，利润为因变量。若现象间只存在相关关系并不存在明显的因果关系，如每万元产值耗电量与产值之间的关系，究竟以哪种现象为自变量，哪种现象为因变量，则要根据研究的目的来决定。

5.1.2 相关关系的分类

社会经济现象之间的相关关系是错综复杂的，表现为各种不同的形态和类型。现从不

同的角度对其进行划分。

1．按相关的变量多少不同，可分为单相关和复相关

- 单相关。两个变量之间的相关关系称为单相关，即一个自变量和一个因变量之间的依存关系。
- 复相关。3 个或 3 个以上变量之间的相关关系称为复相关或多元相关，即一个因变量和 2 个自变量之间的复杂依存关系。

2．按相关的表现形式不同，可分为线性相关和非线性相关

- 线性相关（直线相关）。当相关关系的一个变量变动时，另一个变量也相应地发生大致均等的变动，这种相关关系称为线性相关。
- 非线性相关（曲线相关）。当相关关系的一个变量变动时，另一个变量也相应地发生变动，但这种变动是不均等的，这种相关关系就称为非线性相关。

3．按相关的程度不同，可分为不相关、完全相关和不完全相关

- 不相关。如果两个变量彼此的数量变化互相独立，这种关系为不相关。
- 完全相关。如果一个变量的数量变化由另一个变量的数量变化所唯一确定，这时两个变量间的关系称为完全相关。这种情况下，相关关系实际上是函数关系。所以，函数关系是相关关系的一种特殊情况。
- 不完全相关。如果两个变量之间的关系介于不相关和完全相关之间，称为不完全相关。大多数相关关系属不完全相关。

4．按相关的方向不同，可把线性相关分为正相关和负相关

- 正相关。是指两个变量的变化方向一致，都是增长趋势或下降趋势，如图 5-1 所示。
- 负相关。指两个变量变化趋势相反，一个下降而另一个上升，或一个上升而另一个下降，如图 5-2 所示。

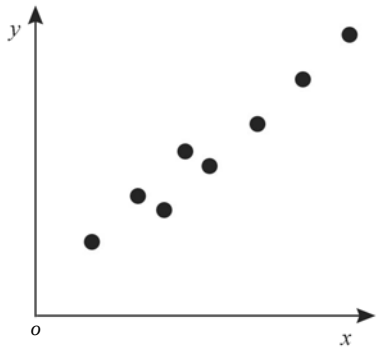


图 5-1　正相关

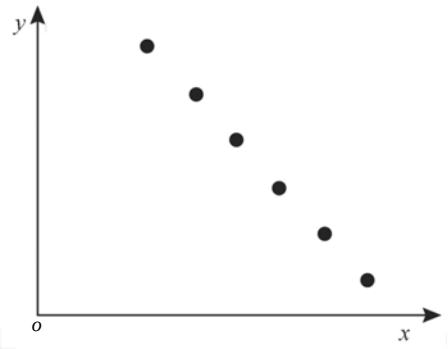


图 5-2　负相关

5.1.3 相关分析的主要内容

相关分析的主要目的，是要对现象间的相互关系的密切程度和变化规律，取得具体数量上的认识，并进一步确定出相互关系的模式，以便进行统计预测和推算。依据对现象分析研究的层次不同，相关分析的内容可分为狭义的相关分析与广义的相关分析。

狭义相关分析的主要内容是：定性与定量分析相结合，正确选择变量，确定变量之间有无相关关系，并确定相关关系的表现形式、密切程度和方向等。

广义的相关分析除了狭义相关分析内容外，还包括：对具有相关关系的变量建立它们之间的数学模型（或称回归方程），并对建立的回归方程及其参数进行显著性检验，如通过检验，可用建立的回归方程来计算；当自变量值有一定变化时，因变量值如何变化，以便揭示变量之间依存关系数量上的规律性。同时，通过测定因变量估计值与实际观测值之间的差异程度，可认识因变量估计值的代表性大小。

5.2 相关关系的测定

5.2.1 相关关系的一般判断

1．定性分析

要分析说明现象之间相关关系的具体数量表现，首先要根据对客观事物的定性认识来判断。由于任何事物都有其质的规定性，质的规定性表明了事物自身与其他事物的区别和联系，因此对事物这种质的规定性的认识和分析，称为定性分析。按照人们认识事物的顺序，先有对事物的定性判断，才能据此进行量的分析。

在对客观现象进行定性分析时，需要借助于有关的经济理论、专业知识和实践经验等。如判明了它们之间没有什么关系，就不需进行相关分析了。可见，对现象之间有无相关关系做出定性判断，是相关分析的一项重要工作。

2．相关表

在对现象之间的相关关系做出定性分析后，再将现象之间的相关关系以表格形式反映，这种表称为相关表。相关表中，以 x 为自变量，y 为因变量，每个自变量都有它相对应的因变量，并在表中一一对应排列。通过相关表可初步看出相关关系的形式、密切程度和相关方向。

➪【例 5-1】

现在以某地区 10 个同类工业企业 2004 年的有关生产性固定资产价值与工业增加值的资料，编制相关表，如表 5-1 所示。

表 5-1　生产性固定资产值与工业增加值相关表

企业编号	生产性固定资产价值 x（百万元）	工业增加值 y（百万元）
1	2.9	19.2
2	2.9	21.1
3	5.0	25.1
4	6.0	30.2
5	6.0	33.8
6	7.0	37.5
7	8.0	37.5
8	9.0	40.0
9	9.0	43.0
10	10.0	50.1

表 5-1 称为简单相关表，它是在原始资料较少时适合使用的一种形式。当原始资料较多时，需要编制分组相关表。分组相关表与变量数列编制方法相同，这里从略。从表中可看出，随着生产性固定资产增多，工业增加值有增加的趋势。

3．相关图

将现象之间的相关关系，通过图像来表示，这种图像称为相关图，也称为散点图。相关图的绘制是在直角坐标中，以横轴表示自变量，纵轴表示因变量，标出每对变量值的坐标点或散布点，表示其分布的状况。通过相关图，可以大致看出两个变量之间有无相关关系及相关的形态、方向和密切程度。现在用表 5-1 的资料绘制相关图，如图 5-3 所示。

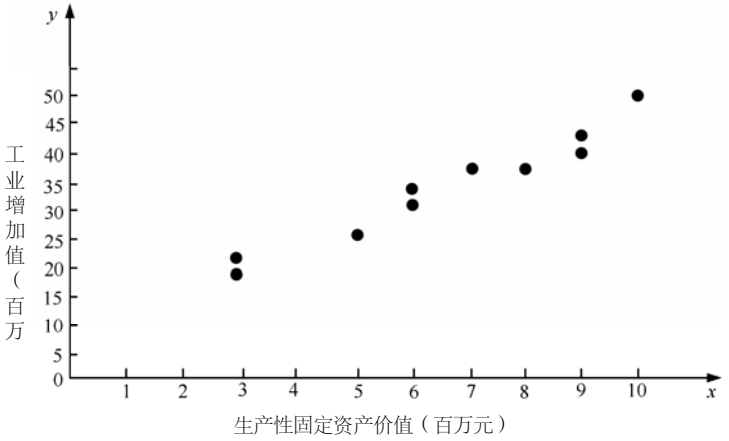


图 5-3　相关图

从图 5-3 中生产性固定资产价值与工业增加值的 10 个散布点来看，因变量 y 随自变量 x 的增大而增大、减少而减少，并且散布点的分布近似地表现为一条直线。由此可以判断生产性固定资产价值与工业增加值这两个变量之间存在着直线正相关关系。

5.2.2 相关系数

1．相关系数的意义

相关系数是在直线相关条件下，说明两个变量之间相关关系密切程度的统计分析指标。相关系数通常用 r 表示。

相关系数比相关表和相关图更能概括表现相关的形式和程度。根据相关系数的大小，或把若干个相关系数加以比较，可以发现现象在发展变化中具有决定作用的因素，因而相关系数对于判断变量之间相关关系的密切程度，有其重要的意义。

相关系数的取值范围是在–1 和+1 之间，即$-1 \leqslant r \leqslant 1$。当 $r>0$，为正相关，$r<0$ 则为负相关。相关系数 r 的绝对值越接近于 1，表示相关关系越强；越接近于 0，表示相关关系越弱。如果$|r|=1$，则表示两个变量完全直线相关。如果$|r|=0$，则表示两个变量完全不相关。但要注意的是，r 只表示 x 与 y 的直线相关密切程度，当 r 很小甚至为 0 时，并不表示 x 与 y 之间不存在其他非直线型的相关关系。

为在实际分析时有个判断标准，现将相关关系密切的等级列于表 5-2 中。

表 5-2　相关密切等级表

相关系数绝对值$\lvert r \rvert$	相关密切程度等级
0.3 以下	不相关
0.3～0.5	低度相关
0.5～0.8	显著相关
0.8 以上	高度相关

2．相关系数的计算

相关系数的测定方法，直接来源于数理统计中相关系数的定义。这里，我们不对公式做理论上的推导和证明，只简要介绍各公式间的联系关系。相关系数的定义公式为：

$$r=\frac{\sigma_{xy}^2}{\sigma_x\sigma_y}=\frac{\frac{1}{n}\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{1}{n}\sum(x-\bar{x})^2}\sqrt{\frac{1}{n}\sum(y-\bar{y})^2}} \tag{5-1}$$

式中，n 是资料项数；σ_x 是 x 变量的标准差；σ_y 是 y 变量的标准差；σ_{xy}^2 是两个变量的协方差。



式（5-1）的分子分母中，有公因子$1/n$，同时约掉化简得：

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} \tag{5-2}$$

式（5-2）是通过各变量离差乘积的方法来计算相关系数的，也称为“积差法”相关系数公式。由于各变量的离差通常带有小数，因而“积差法”公式的计算结果往往缺乏准确性。在实际应用中，如果依据原始资料计算相关系数，可用式（5-2）的简化公式计算，见式（5-3）。

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^2-(\sum x)^2}\sqrt{n\sum y^2-(\sum y)^2}} \tag{5-3}$$

该式是由积差法公式推导出来的。按照该公式计算系数，只需列 3 个计算栏：xy，x^2，y^2。避免了平均数、协方差、标准差的直接计算，大大简化了运算过程。

➪【例 5-2】

现在用表 5-1 的资料，计算 8 个企业月产量与生产费用之间的相关系数。所需资料如表 5-3 所示。

表 5-3　相关系数计算表

企业序号	月产量 x（千吨）	生产费用 y（万元）	x^2	y^2	xy
1	1.2	62	1.44	3 844	74.4
2	2.0	86	4.00	7 396	172
3	3.1	80	9.61	6 400	248
4	3.8	110	14.44	12 100	418
5	5.0	115	25.00	13 225	575
6	6.1	132	37.21	17 424	805.2
7	7.2	135	51.84	18 225	972
8	8.0	160	64.00	25 600	1 280
合计	36.4	880	207.54	104 214	4 544.6

由表中资料得如下有关数据：

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^2-(\sum x)^2}\sqrt{n\sum y^2-(\sum y)^2}}=\frac{8\times 4\,544.6-36.4\times 880}{\sqrt{8\times 207.54-36.4^2}\sqrt{8\times 104\,214-880^2}}$$
$$=0.969\,7$$

计算结果表明生产性固定资产价值和工业增加值之间存在高度的正相关关系。

通过以上的计算与分析过程，我们看到：统计所研究现象之间的相关关系，应该是真实的、客观存在的联系关系，而不是主观臆造或形式上的偶然巧合。这就要求我们在实际进行相关关系分析时，应依据有关的科学理论，通过观察和试验，在对现象做深入分析的基础上，来确定这种联系关系，而且还要经过理论和实践的进一步检验。只有这样，才可能得到正确的结论。

5.3 一元线性回归分析

5.3.1 回归分析的概念

相关分析中的相关系数指标可以从数量上说明在直线相关的条件下，变量之间相关关系的方向和密切程度。但它不能反映一个变量发生一定数量的变化时，另一个变量会相应地发生多少变动。为解决这一问题，就必须采用回归分析方法。

回归分析是指对具有相关关系的变量，依据其关系形态，选择一个合适的数学模型，用来近似地表示变量间数量平均变化关系的一种统计方法（或称回归方程式）。回归分析的内容很多，按分析变量的多少不同，可分为一元回归分析与多元回归分析；按分析变量间的表现形态不同，可分为线性回归分析与非线性回归分析等。本节只讨论一元线性回归分析的有关理论和方法。



回归的概念来源于英国统计学家 Francis Galton 在 19 世纪末期研究父母和他们的孩子的身高时所采用的方法。此研究发现，孩子的身高在很大程度上依赖于父母的身高，身材较高的父母他们的孩子也较高，但是孩子比身材很矮或很高的父母更加趋于平均值，这种向中间值靠近的趋势即被称做回归效应。由此发展出来的研究两个变量相关关系的方法即被称为回归分析。不过现如今在几乎所有应用回归分析的情形当中，“回归”的特征并不那么明显，我们更多的是关注因变量随自变量变化而变化的协同关系。

5.3.2 相关分析与回归分析的区别与联系

1. 相关分析与回归分析的区别

- 相关分析所研究的两个变量是对等关系；而回归分析所研究的两个变量不是对等关系，必须根据研究目的，先确定其中一个是自变量，另一个是因变量。



- 对两个变量 x 和 y 来说，相关分析只能计算出一个反映两个变量间相关密切程度的相关系数，计算中改变 x 和 y 的位置不影响相关系数的数值；回归分析中，对互有因果关系的变量，有时可以根据研究目的的不同分别建立两个不同的回归方程。以 x 为自变量，y 为因变量，可以得出 x 对 y 的回归方程；以 y 为自变量，x 为因变量，可得出 x 对 y 的回归方程。
- 相关分析对资料的要求是，两个变量都必须是随机变量；而回归分析对资料的要求是，自变量是可以控制的变量或给定的变量，而因变量是随机变量。

2．相关分析与回归分析的联系

- 相关分析是回归分析的基础和前提。如果缺少相关分析，没有从定性上说明现象间是否具有相关关系，没有对相关关系的密切程度做出判断，就不能进行回归分析，即便进行了回归分析，其回归方程的代表性也是值得质疑的。
- 回归分析是相关分析的深入和继续。对现象间相关分析仅仅说明其具有密切的相关关系是不够的，只有进行了回归分析，拟合了回归方程，才可能进行有关分析的回归预测，相关分析才有实际的意义。因此，如果仅有回归分析而缺少相关分析，将会因为缺乏必要的基础和前提而影响回归分析的可靠性；如果仅有相关分析而缺少回归分析，就会降低相关分析的意义。只有把两者结合起来，才能达到分析研究的目的。

5.3.3　一元线性回归模型

1．一元线性回归模型的描述

一元线性回归模型又称简单直线回归模型，它是分析具有线性相关关系的两个变量之间数量变动关系的数学表达式。

当 x 与 y 变量互为因果关系时，依据分析研究的目的不同，可建立两个直线回归方程式。

（1）y对x的直线回归方程

$$\hat{y} = a + bx \tag{5-4}$$

式中，$\hat{y}$ 是因变量 y 的估计理论值；x 是自变量的实际值；a，b 是待定参数。

a，b 的几何意义是：a 为直线方程的截距，b 为斜率。其经济意义是：a 为当 x 为 0 时 y 的估计值；b 是当 x 每增加一个单位时 y 的平均增加量，b 也称 y 对 x 的回归系数。

（2）x对y的直线回归方程

$$\hat{x} = a' + b'y \tag{5-5}$$

式（5-5）中 a' 与 b' 的意义相同于式（5-4）中的 a 与 b，只是 x 与 y 的位置互换而已。

当 x 与 y 变量只有单向的依存关系时，只能建立一个直线回归方程，一般是 y 对 x 的回归直线，即式（5-4）。

2．一元线性回归模型的参数估计

在相关图中，如果自变量与因变量对应的散布点近似为直线，或计算出的相关系数具有显著的直线相关关系，则可拟合一条回归直线。当然，我们拟合回归直线的目的是要找到一条理想的直线，用直线上的点来代表所有的相关点。统计理论已证明，用最小平方法配合的直线最具有代表性，也称最佳线。

应用最小平方法配合直线，其基本要求是实际值与估计理论值的离差平方和为最小。用公式表示为：

$$\sum (y-\hat{y})^2=\sum(y-a-bx)^2=\text{最小值}$$

$$\text{设}\quad Q=\sum(y-a-bx)^2$$

则 Q 是两个待定参数 a，b 的函数。要使 Q 为最小值，就要用数学中对二元函数求极值的原理，求 Q 关于 a 和 b 的偏导数，并令其等于零，整理得出直线回归方程中求解参数 a 和 b 的标准方程组。

$$\begin{cases}\sum y=na+b\sum x\\ \sum xy=a\sum x+b\sum x^2\end{cases} \tag{5-6}$$

这就是求解参数 a，b 的二元一次方程组。解之即得求 a，b 的公式，如下所示：

$$\begin{cases}b=\dfrac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}\\ a=\dfrac{\sum y}{n}-b\dfrac{\sum x}{n}\end{cases} \tag{5-7}$$

➪【例 5-3】

现仍用表 5-1 某地区的 8 个同类工业企业 2004 年的有关月产量与生产费用的资料为例，具体说明求解 a，b 值和建立直线回归方程的方法。

从相关图 5-3 中看出，生产性固定资产价值与工业增加值为线性相关，并通过表 5-3 相关系数的测算，得知二者之间为高度的正相关关系，由此可建立 y 对 x 的直线回归方程。根据表 5-4 的资料求出所需的有关数据。

表 5-4　直线回归模型计算表　　单位：100 万元

企业序号	月产量 x（千吨）	生产费用 y（万元）	x^2	xy	估计理论值 $\hat{y}$
1	1.2	62	1.44	74.4	66.79
2	2.0	86	4.00	172	77.11

续表

企业序号	月产量 x（千吨）	生产费用 y（万元）	x^2	xy	估计理论值 $\hat{y}$
3	3.1	80	9.61	248	91.3
4	3.8	110	14.44	418	100.33
5	5.0	115	25.00	575	115.81
6	6.1	132	37.21	805.2	130
7	7.2	135	51.84	972	144.19
8	8.0	160	64.00	1 280	154.51
合计	36.4	880	207.54	4 544.6	442.423

$$n=8 \quad \sum x=36.4 \quad \sum x^2=207.54 \quad (\sum x)^2=1\,324.96$$
$$\sum y=880 \qquad \sum xy=4\,544.6$$

将上述数据代入式（5-7），求解 a 和 b 的数值。

$$b=\frac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-\sum(x^2)}=\frac{8\times 4\,544.6-36.4\times 880}{8\times 207.54-1\,324.96}$$
$$=12.90$$
$$a=\frac{\sum y}{n}-b\frac{\sum x}{n}=\frac{880}{8}-12.90\times\frac{36.4}{8}=51.31$$

将 a 和 b 的数值代入式（5-4），得出工业增加值对生产性固定资产价值的直线回归方程：

$$\hat{y}=51.31+12.90x$$

回归方程的实际意义是：a=51.31，是回归直线在 y 轴上的截距；b=12.9，表示月产量每增加一个单位（千吨），生产费用平均增加 12.9（万元）。

回归直线确定后，将各企业的月产量依次代入方程式，即可求得各企业生产费用的理论值，将其填入表 5-4 中，以供绘制回归直线图和计算回归估计标准误差使用。例如：

企业 1　$\hat{y}=51.31+12.9\times 1.2=66.79$（万元）

企业 2　$\hat{y}=51.31+12.9\times 2.0=77.11$（万元）

利用回归直线模型还可以进行预测。例如，某企业月产量预计可达 4 000 吨，在其他条件相对稳定时，可预测其生产费用为：

$$\hat{y}=51.31+12.9\times 4=102.91\text{（万元）}$$

这里需要指出，一个直线回归方程只能做一种推算，不能进行另一种相反推算。即只能以自变量 x 推算因变量 y，而不能以 y 推算 x。如上例所配合的直线回归方程，只能在给定生产性固定资产价值时，来推算工业增加值，不可给定工业增加值来推算固定资产价值。而在互为因果关系的变量之间，根据研究需要，可建立 y 对 x 的直线回归方程和 x 对 y 的直线回归方程，但此时的两个回归方程是两条不同的回归直线，具有不同的斜率和意义，只能给定自变量推算因变量。

3．回归系数与相关系数的关系

回归系数 b 与相关系数 r 有着非常密切的数量关系，在相关分析和回归分析中二者之间可互相推算。

又根据回归系数式（5-7），得

$$r=\frac{\sigma_{xy}^2}{\sigma_x\sigma_y}$$

根据相关系数式（5-1）

$$b=\frac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}=\frac{\sigma_{xy}^2}{\sigma_x^2}$$

得

$$\frac{b}{r}=\frac{\sigma_y}{\sigma_x}$$

所以

$$b=r\frac{\sigma_y}{\sigma_x} \qquad (5\text{-}8)$$

由于 σ_y/σ_x 为非负数，因此，由式（5-8）可知 b 与 r 同号且 b 正比于 r，这是一个很重要的结论。它为在某场合下依据 r 来推算 y 随 x 的变动情况提供了理论依据。

现在用以上有关算例的结果来验证二者的关系。由表 5-3 的结果及进一步计算得

$$\sigma_x=2.331\ 24，\ r=0.97，\ \sigma_y=30.44$$

再由表 5-4 的结果得

$$b=4.154\ 4$$

则

$$b=r\sigma_y/\sigma_x=0.991\ 8\times 9.769\ 8/2.332\ 4=4.154\ 4$$

$$r=b\sigma_y/\sigma_x=4.154\ 4\times 2.332\ 4/9.769\ 8=0.991\ 8$$

可见，推算结果与依据原始资料计算的结论完全一致。

5.3.4　估计标准误差

根据所配合的回归直线方程式推算出各个估计值 $\hat{y}$ 后，其估计的可靠性如何，则要通过计算估计标准误差来判断。在表 5-4 中，因变量的估计值 $\hat{y}$ 与实际观察值 y 是不等的，其估计误差为 $y-\hat{y}$。估计误差的大小反映着估计值的准确性。但我们要观察的不是某一个变量值与估计值的误差，而是整体的差别情况，这就需要用估计标准误差指标来反映。

1．估计标准误差的意义与计算

估计标准误差是用来说明回归方程代表性大小的统计分析指标。其计算原理与标准差基本相同，估计标准误差说明估计理论值的代表性。若估计标准误差小，表明回归方程估计准确程度高，代表性大；反之，则估计准确程度低，代表性小。估计标准误差的公式如下。

（1）定义公式

定义公式反映了实际观察值 y 与估计值 $\hat{y}$ 之间的平均离差程度。

$$S_y=\sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} \tag{5-9}$$

式中，S_y 是估计标准误差；$n-2$ 是自由度，因在一元线性回归方程中，计算了 a 和 b 两个参数，即失去了两个自由度。

按照上面的定义公式计算估计标准差十分烦琐，运算较大，因为它需要计算出因变量 y 所有的估计值。实际中，运用简化公式计算估计标准误差。

（2）简化公式

在已知直线回归方程下，简化公式是根据回归直线方程中的参数 a 和 b 计算的，特点是能与相关分析和回归分析中的数据资料相结合，其计算结果等同于定义公式。

$$S_y=\sqrt{\frac{\sum y^2-a\sum y-b\sum xy}{n-2}} \tag{5-10}$$

➪【例 5-4】

现仍用表 5-4 的有关资料，说明回归标准误差的计算方法。

经过有关计算已知

$$\sum(y-\hat{y})^2=442.43, n-2=6, \sum y^2=104\,214, \sum y=880, \sum xy=4\,544.6, a=51.31, b=12.9$$

按照定义式（5-9）得

$$S_y=\sqrt{\frac{\sum(x-\hat{y})^2}{n-2}}=\sqrt{\frac{442.43}{6}}=8.59(\text{万元})$$

按照简化式（5-10）得

$$S_y=\sqrt{\frac{\sum y^2-a\sum y-b\sum xy}{n-2}}$$
$$=\sqrt{\frac{104214-51.31\times 880-12.9\times 4544.6}{6}}$$
$$=8.59(\text{万元})$$

其结果表明，8 个企业生产费用的估计理论值与实际值的平均误差为 8.59 万元。由此可见，只有把回归估计值与估计的标准误差结合起来分析运用，才更具有意义。

2．估计标准误差与相关系数的关系

估计标准误差 S_y 与相关系数 r 在数量上也存在着密切的关系，这就从另一个角度说明了相关分析与回归分析之间的联系。二者之间的关系可由下列公式表述：

$$r=\sqrt{1-\frac{S_y^2}{\sigma_y^2}} \qquad (5\text{-}11)$$

$$S_y=\sigma_y\sqrt{1-r^2} \qquad (5\text{-}12)$$

式（5-11）、式（5-12）中有关符号的意义同前。

式（5-11）和式（5-12）的意义在于：从互相联系的两个算式中可以看出 r 与 S_y 的变化方向是相反的。当 r 越大时，S_y 越小，这时相关密切程度较高，回归直线的代表性就大；$r=\pm 1$ 时，$S_y=0$，现象间完全相关，各相关点均落在回归直线上，此时对 x 的任何变化，y 总有一个相应的值与之对应。当 r 越小时，S_y 就越大，这时相关密切程度则较低，回归直线的代表性就小。$r=0$ 时，S_y 取得最大值，这时，现象间不存在直线关系。

在实际的相关分析中，一般不常用这种方法计算相关系数。这种计算方法存在两个缺陷：

- 需要先求出回归直线方程，计算出估计标准误差，才能求得相关系数。而从一般的认识程序来看，只有相关关系密切的情况下，配合回归方程才有意义；如果关系不密切，下一步的计算就不必要了，因而要求先计算相关系数以判断相关关系的密切程度。
- 以这种方法计算出的结果，难以判断是正相关还是负相关。

所以实际工作中常常采用另一种推算方法，即根据式（5-12）由相关系数 r 去推算估计标准误差 S_y。

5.4 应用相关分析与回归分析应注意的问题

5.4.1 在定性分析的基础上进行定量分析

在定性分析的基础上进行定量分析，是保证正确运用相关分析和回归分析的必要条件。也就是在确定哪些变量是自变量，哪些变量是因变量之前，必须对所研究的现象有充分的认识。相关分析的方法不能解释相关关系产生的原因，它本身不能判断现象之间是否存在质的联系。这些问题的解决，必须依靠对现象的定性分析。即须根据社会经济理论、专业知识、实际经验对事物进行分析来判定。如果把本来没有内在关系的现象进行相关分析，将会导致虚假相关的错误。比如，研究某地区人口出生数量与该地区老年人口再婚数量之间的关系，显然是没有任何意义的。若据此进行回归分析推算预测，其结论也是荒谬的。

5.4.2 要注意现象质的界限及相关关系作用的范围

在进行相关分析和回归分析时，要注意现象质的界限及相关关系作用的范围。超出了这个范围，分析的结果就会有悖于客观实际。我们用数学模型得到的回归方程，一般都是根据一定范围内的有限资料来计算的，其有效性，一般只适用于该范围内，不适宜用于该范围外。也就是说，利用回归方程，一般只适宜用于内插预测，不适宜用于外推预测。这是由于最小平方法指的是对现有资料范围配合一条最佳线，如果外推到范围以外，就不一定是最佳线了。我们根据样本资料所配合的回归方程是在一定条件下建立的，因此也只能在一定条件下应用。如果忽视了这些条件，把这种关系无限制地向外推广是不正确的，由此得到的结论也是值得质疑的。例如，农作物生产量与施肥量只在一定的范围内才具有正相关关系，施肥量超过一定的限度，产量不但不会增加，反而会减少。因此，在用相关分析、回归分析进行推算预测时，要注意它的作用范围。

5.4.3 要将各种分析指标结合运用

相关分析与回归分析中的各种分析指标，既有区别又有联系。相关分析的相关系数，是反映现象之间直线关系密切程度的指标；回归分析的直线回归方程，反映着现象之间数量变化的联系；回归估计标准误差则表明了回归直线的代表性大小。它们分别从不同的侧面揭示了现象之间数量联系上的特征。因而在相关分析、回归分析时，只有把几者有机地结合起来，才能更加准确地描述相关现象之间数量变化的规律性。

5.4.4 要对参数的有效性进行检验

在回归分析中，要对回归模型计算出来的参数（包括常数项及回归系数）的有效性进行显著性检验，以判断回归预测的有效性。经过检验，如发现某回归系数的数值没有显著

意义，或某些自变量间存在着多重共线性，则具有这种情况的自变量应从回归方程中剔除，以保证回归预测的有效性和准确性。关于这类问题的检验方法本章不做介绍，读者可参阅有关相关分析与回归分析方面的著作。

本章小结

1. 相关和回归是对客观事物的依存关系定量分析的科学方法，它们之间既有联系又有区别。相关分析着重于随机变量之间相关形式和相关程度的分析；回归分析着重于变量之间具体变动关系的分析，通过建立回归模型，控制或给定自变量对因变量进行评估和预测。

2. 构建一元线性回归模型，掌握两个参数的计算公式以及回归系数的统计意义。回归系数的符号同相关系数符号一致，表明变量变化的方向和变量变动比率

3. 对一元线性回归方程的可靠程度进行判断。估计标准误差的计算公式和表达实际意义。一般情况下，应用一元线性回归模型估计和预测，只限于模型赖以建立的资料本身的范围内；如果用于外推估计，必须十分谨慎。

4. 必须把定量分析同定性分析紧密结合起来，以定性分析为基础，确定自变量与因变量。对于客观事物是否存在联系、具体关系认定以及回归模型计算结果的解释等，都必须运用科学理论、专业知识和实践经验进行定性分析才能得出正确判断。

复习思考题

一、概念题

函数关系　相关关系　相关系数　相关分析　回归分析　自变量　因变量　回归方程　回归系数　估计标准误差

二、简答题

1. 什么是函数关系？什么是相关关系？二者有什么区别与联系？
2. 什么是正相关、负相关、不相关？试举例说明。
3. 相关系数的意义是什么？怎样利用相关系数来判别现象之间的相关关系？
4. 回归分析与相关分析的联系和区别有哪些？
5. 配合回归直线方程有什么要求？方程中参数 a，b 的经济意义是什么？
6. 什么叫估计标准误差？简述估计标准误差在回归与相关分析中的意义。
7. 应用相关分析与回归分析应注意哪些问题？

三、练习题

1．判断题（把“√”或“×”填在题后的括号里）

（1）现象之间的相关关系可以用一个严格的数学表达式表示出来。（　　）

（2）相关系数为+1 时，两变量完全正相关；相关系数为–1 时，两变量完全负相关。（　　）

（3）当相关系数等于零时，两变量之间不存在相关关系。（　　）

（4）若变量 x 的值增加时，变量 y 的值也增加，说明 x 和 y 之间存在正相关关系；若变量 x 的值减少时，变量 y 的值也增加，说明 x 和 y 之间存在负相关关系（　　）

（5）回归系数和相关系数都可以用来判断现象之间相关的密切程度。（　　）

（6）利用最小平方法配合的直线回归方程，要求实际测定的所有相关点和直线上的点的距离平方和为零。（　　）

2．单选题

（1）现象之间相关关系的类型有（　　）。

A．相关关系和因果关系　　B．相关关系和随机关系

C．相关关系和函数关系　　D．函数关系和因果关系

（2）存在相关关系的两变量间关系的性质（　　）。

A．因果关系

B．共变关系

C．可能是因果关系，也可能是共变关系

D．以共变形式表现的因果关系

（3）当自变量的数值确定后，因变量的数值布固定，这种关系属于（　　）。

A．相关关系　　B．函数关系　　C．回归关系　　D．因果关系

（4）在相关分析中，要求相关的两变量（　　）。

A．都是随机变量　　B．都不是随机变量

C．因变量是随机变量　　D．自变量是随机变量

（5）降雨量与农作物亩产量之间的关系是（　　）。

A．当自变量的值变动时，因变量的值也随着变动

B．当自变量的值变动时，因变量的值不随着变动

C．当自变量的值增加时，因变量的值也随着变动

D．当自变量的值增加时，因变量的值反而减少

3．多选题

（1）相关关系与函数关系的数量关系为（　　）。

A．函数反映的是确定性的数量关系

B．相关关系表明的是现象间非确定性的数量关系

C．相关关系表现为变量间完全确定的依存关系

D. 函数关系可以为变量间不完全确定的依存关系
E. 相关关系不能用数学函数式表达

（2）下列属于正相关的现象有（　　）。
A. 商品销售额随着商品销售量的增加而增加
B. 某产品产量随工人劳动生产率的提高而增加
C. 流通费用率随商品销售额的增加而减少
D. 单位成本随着单位产品原材料消耗的降低而减少
E. 产品产量随工人劳动生产率的提高而增加

（3）下列属于负相关的现象有（　　）。
A. 商品流转的规模越大，流通费用水平越低
B. 流通费用率随商品销售额的增加而减少
C. 施肥量增加，粮食平均亩产量也增加
D. 生产单位产品所耗工时随劳动生产率的提高而减少
E. 产品产量随工人劳动生产率的提高而增加

（4）判断现象之间有无相关关系的方法有（　　）。
A. 对客观现象作定性分析　　B. 编制相关表
C. 计算相关系数　　D. 计算估计标准误差
E. 绘制相关图

（5）变量间相关关系按其相关密切程度划分有（　　）。
A. 不相关　　B. 不完全相关　　C. 完全相关　　D. 线性相关
E. 非线性相关

（6）直线相关分析的特点是（　　）。
A. 两个变量是对等关系　　B. 只能计算出一个相关系数
C. 相关系数有正负号　　D. 相关的两个变量必须都是随机的
E. 相关系数的大小反映两个变量之间相关的密切程度

4．计算题

（1）2001—2008年人均收入和耐用消费品销售额资料如表5-5所示。根据表5-5资料：

表5-5　2001—2008年人均收入和耐用消费品销售额资料　　单位：万元

年份	人均收入	耐用消费品销售额
2001	3.0	80
2002	3.2	82
2003	3.4	85

续表

年份	人均收入	耐用消费品销售额
2004	3.5	90
2005	3.8	100
2006	4.0	120
2007	4.5	140
2008	5.2	145

1）绘制相关散点图，并观察相关关系的趋势。

2）计算相关系数，并判别相关关系的密切程度和方向。

（2）某工业企业上半年产品产量与单位成本资料如表 5-6 所示。根据表 5-6 资料：

表 5-6　某工业企业上半年产品产量与单位成本资料

月　份	产品产量（千件）	单位成本（元 / 件）
1	2	73
2	3	72
3	4	71
4	3	73
5	4	69
6	5	68

1）计算相关系数，判断其相关方向和程度。

2）建立回归直线方程（单位成本为因变量），指出产量每增加 1000 件时单位成本平均下降多少。

（3）某高校教育经费与学生人数连续 6 年的统计资料如表 5-7 所示。根据表 5-7 资料：

表 5-7　某高校教育经费与学生人数的统计资料

教育经费（万元）	高校学生人数（万人）
316	11
343	16
373	18
393	20
418	22
455	25

1）建立直线回归方程，估计教育经费为 500 万元的在校学生数。

2）计算估计标准误差。

（4）试根据 $\sigma_x^2 = 25$，$\sigma_y^2 = 36$，$r = 0.9$，$a = 2.8$，求：

1）构建直线回归方程。

2）计算估计标准误差 S_y。

CHAPTER 6

第 6 章　时间数列分析

章首引例

时间数列将某一统计指标在不同时间上的数值按时间先后顺序排列而形成的一种统计数列。例如，表 6-1 根据《江苏省统计年鉴》列举 2003—2007 年江苏省几个主要指标时间数列。

表 6-1　江苏省 2003—2007 年国民经济和社会某些主要指标的时间数列 ①

年　份	2003	2004	2005	2006	2007
地区生产总值（亿元）	12 460.83	15 512.35	18 305.66	21 645.08	25 741.15
第二产业产值（亿元）	6 787.11	8 770.32	10 355.04	12 250.84	14 306.40
第二产业占地区生产总值比重（%）	54.47	56.54	56.57	56.60	55.58
年末全省就业人员（万人）	4 468.67	4 482.52	4 510.12	4 564.76	4 618.14
全省在岗职工年平均工资（元）	15 712	18 202	20 957	23 782	27 374

由于社会经济现象随时都在变化，必须随时掌握其变化的过程和发展趋势，这就需要我们研究和分析经济现象随时间的变迁而变化的规律，并据以预测未来时间的发展趋势值提供依据。

本章学习目标

- ☑ 理解时间数列的概念、作用、种类和编制原则；
- ☑ 熟练掌握时间数列各种水平指标和速度指标计算方法与应用；
- ☑ 掌握时间数列变动趋势和季节变动规律的分析意义与方法。

6.1 时间数列的编制

6.1.1 时间数列的意义

时间数列又称时间序列、动态数列，它是将某一统计指标在不同时间上的数值按时间

① 摘自《江苏省统计年鉴》，www.jssb.gov.cn/jstj/jsnj。

先后顺序排列而形成的一种统计数列。

从表 6-1 可以看出，时间数列由两个基本要素构成：一个是现象所属的时间；另一个是反映现象的统计指标数值，也称为发展水平。

时间数列在统计研究中具有重要作用，主要表现在以下几个方面：

- 通过时间数列可以描述某一社会经济现象的发展状况、变化的过程和规律。
- 利用时间数列资料便于计算一系列动态分析指标。
- 根据时间数列可以揭示现象发展变化的趋势，从而为统计预测提供依据。

6.1.2 时间数列的种类

时间数列按其指标表现形式的不同，可分为绝对数时间数列、相对数时间数列和平均数时间数列三种。其中，绝对数时间数列是基本数列，相对数时间数列和平均数时间数列是在其基础上派生的。

1．绝对数时间数列

总量指标是反映一定时间、地点和条件下某种现象总体规模或水平的统计指标，它的表现形式为绝对数。把同一总量指标在不同时间上的数值按时间先后顺序排列而形成的时间数列称为绝对数时间数列。它用来反映被研究现象在各期达到的绝对水平及其发展变化情况。它又可分为时期数列和时点数列。

（1）时期数列

在绝对数时间数列中，如果各项指标都是反映某种现象在一段时期内发展过程的总量，即时期指标，这种绝对数时间数列就称为时期数列，如表 6-1 中的不同时期地区生产总值以及不同年份第二产业增加值都是时期数列。时期数列的特点是：

- 数列中各个指标的数值是可以相加的，即相加具有一定的经济意义。由于时期数列中每个指标的数值是表示在一段时期内发展过程的总量，所以相加后的数值就表示现象在更长一段时期内发展过程的总量。
- 数列中每一个指标数值的大小与所属的时期长短有直接的联系。在时期数列中，每个指标值所包括的时间长度，称为“时期”。时期的长短，主要根据研究目的而定，可以是一日、一旬、一月、一季、一年或更长时期。一般来说，时期越长，指标数值就越大，反之就越小。
- 数列中每个指标的数值，通常是通过连续不断地登记而取得的。

（2）时点数列

在绝对数时间数列中，如果各项指标都是反映现象在某一时刻上（瞬间）的总量，即时点指标，这种绝对数时间数列就称为时点数列，如表 6-1 的年末江苏省就业人员就是个时点数列。时点数列有如下特点：



- 数列中各个指标的数值是不能相加的，相加不具有实际经济意义。这是由于时点数列中每个指标数值都是表明现象在某一瞬间上的数量，相加以后无法说明属于哪一时点的数量。
- 数列中每一个指标数值的大小与其时间间隔长短没有直接联系。在时点数列中两个相邻指标数值在时间上的距离叫做“间隔”。由于时点数列的每个指标数值只表明现象某一时点上的数量，年末数值可能大于月末数值，也可能小于月末数值，因此，它的指标数值大小与时间间隔长短没有直接联系。
- 数列中每个指标的数值，通常是通过一定时期登记一次而取得的。

2．相对数时间数列

相对指标是说明现象之间数量对比关系的指标，用两个或两个以上有联系的指标数值对比来求得，其结果表现为相对数。把同一相对指标在不同时间上的数值按时间先后顺序排列而形成的时间数列称为相对数时间数列。它反映社会经济现象之间相互联系的发展过程。表 6-1 中的第二产业增加值占地区生产总值的比重就是一个相对数时间数列。在相对数时间数列中，各个指标数值是不能相加的。

3．平均数时间数列

把同一平均指标在不同时间上的数值按时间先后顺序排列而形成的时间数列称为平均数时间数列。它反映现象在一段时间内一般水平发展变化的过程。表 6-1 中的全省在岗职工年平均工资就是一个平均数时间数列。在平均数时间数列中，各个指标数值也是不能相加的，相加没有经济意义。

6.1.3　编制时间数列的原则

编制时间数列的目的是通过同一指标不同时间的数值对比来反映社会经济现象的发展变化及其规律性。因此，保证数列中各个指标数值之间的可比性，就成为编制时间数列应遵守的基本原则。具体来说，应注意下列四点。

1．时期长短应该统一

在时期数列中，由于各个指标数值的大小与时期长短有直接的关系，因此，各个指标所属的时期长短应当前后统一。时间越长，指标数值就越大，反之就越小。时期长短不一，往往就很难进行直接比较。但这个原则也不能绝对化，有时为了特殊的研究目的，也可将时期不等的指标编成时期数列，如表 6-2 所示。

表 6-2　我国不同历史时期的钢产量

年份	1900—1948	1953—1957	1991—1995
钢产量（万吨）	760	1 667	42 708

从表 6-2 中可以明显看出，我国第一个五年计划时期的钢产量超过旧中国将近半个世纪钢产量 1 倍以上，第八个五年计划时期又比第一个五年计划时期有了更快的发展，增长了 24 倍以上。

对于时点数列来说，由于各个指标只反映现象在某一时点的状态，所以不存在时期长短应该统一的问题，两时点间隔长短，对时点指标数值的大小没有直接影响，但为了更有利于对比，时点间隔最好能保持一致。

2．总体范围应该统一

时间数列中，各个指标所包括总体范围前后应该一致。如研究某地区工业生产发展情况，如果该地区的行政区划有了变动，则前后指标就不能直接对比，必须将资料进行适当的调整，以求总体范围的统一，然后再做动态分析。

3．计算方法应该统一

时间数列各项指标的计算口径、计量单位和计算方法应该一致，保持不变。例如，要研究企业劳动生产率的变动，产量用实物量还是用价值量，人数用全部职工数还是用生产工人数，前后都要统一起来。再如，要对比不同时期的工农业产值，就应该注意价格水平的变化，采用统一的不变价格表示。价格标准不同，就不能从指标的对比中正确反映工农业产量的实际变化程度。

4．经济内容要统一

有时时间数列的指标在名称上是一个指标，但经济内容或经济含义不同或有了改变，这也是不可比的。例如，工业企业里的工资指标，按费用要素分组的工资包括全部职工的工资；而按成本项目分组的工资只包括基本生产工人的工资，如果把这样一些指标数值编成时间数列反映现象的变动，就会得出错误的结论。

6.2 时间数列的水平指标

为了进一步分析和认识现象发展变化的结果，根据时间数列可以计算一系列动态分析指标，分水平指标与速度指标两大类。本节主要介绍水平指标，包括发展水平、平均发展水平、增长量和平均增长量。

6.2.1 发展水平

发展水平是时间数列中的每个指标数值，它反映现象在不同时间发展达到的规模和水平，是计算其他动态分析指标的基础。发展水平可表现为绝对数，如工资总额、工业总产值、年末职工人数；也可表现为相对数或平均数，如人口出生率、工人劳动生产率等。

在时间数列中，由于发展水平所处的位置不同，有最初水平、中间水平和最末水平之



分。通常用符号 a 表示发展水平。在时间数列 a_0，a_1，a_2，…，a_{n-1}，a_n 中，a_0 为最初水平，a_n 为最末水平，其余 a_1，a_2，…，a_{n-1} 为中间水平。在对两个时间的发展水平做动态对比时，作为对比基础时期的发展水平称为基期水平，作为研究时期的指标水平称为报告期水平或计算期水平。

6.2.2　平均发展水平

平均发展水平是时间数列中各个指标数值加以平均所得的平均数，在统计上又称为序时平均数或动态平均数，它表明现象在一段时间内发展过程达到的一般水平。

序时平均数与一般平均数（静态平均数）既有共同之处，也有区别。共同之处是二者都是将现象的个别数量差异抽象化了，概括出了现象在数量上达到的一般水平。两者的区别在于：

- 序时平均数是根据时间数列计算的，而一般平均数是根据变量数列计算的；
- 序时平均数所平均的是被研究现象本身的数量在不同时间上的差异，而一般平均数所平均的是总体各单位某一数量标志值的差异；
- 序时平均数是从动态上表明被研究现象本身在一段时期内的平均发展水平，而一般平均数是从静态上说明总体各单位某个数量标志值的平均水平。序时平均数在动态分析中被广泛地运用。

序时平均数可根据绝对数时间数列计算，也可根据相对数和平均数时间数列计算。其中，绝对数时间数列序时平均数的计算方法是最基本的方法。

1. 由绝对数时间数列计算序时平均数

由于绝对数时间数列分时期数列和时点数列，它们各具有不同特点，因而计算序时平均数的方法也就不一样，现分别加以说明。

（1）由时期数列计算序时平均数

由于数列中各项指标数值相加等于全部时期的总量，因此可直接采用简单算术平均法，即将数列中各指标数值相加，再除以时期数求得序时平均数。其计算公式如下：

$$\bar{a}=\frac{a_1+a_2+\cdots+a_{n-1}+a_n}{n}=\frac{\sum a}{n} \qquad (6\text{-}1)$$

式中，$\bar{a}$ 是序时平均数；a 是各期发展水平；n 是时期项数。

（2）由时点数列计算序时平均数

时点数列分为连续时点数列和间断时点数列，它们的序时平均数计算方法因掌握资料的情况不同而异。

1）根据连续时点数列计算序时平均数

如果时点数列的资料是逐日记录，而又是逐日排列的，则称为连续时点数列。

在掌握间隔相等的连续时点（如每日的时点）资料时，用式（6-1）计算求得。

【例 6-1】

江苏省 2003—2007 年各年地区生产总值资料如表 6-1 所示，求年平均地区生产总值。

年平均地区生产总值为：

$$\bar{a} = \frac{12460.83 + 15512.35 + 18305.66 + 21645.08 + 25741.15}{5}$$

$$= \frac{93665.07}{5} = 18733.01(\text{亿元})$$

例如，已知某企业一个月内每天的工人人数，要计算该月内每天平均工人人数，可将每天的工人人数相加，除以该月的日历日数即得。

在掌握间隔不相等的连续时点资料时，如被研究现象不是逐日变动，而是间隔几天变动一次，可用加权算术平均法计算序时平均数。其计算公式为：

$$\bar{a} = \frac{\sum a f}{\sum f} \qquad (6\text{-}2)$$

式中，$\bar{a}$ 是序时平均数；a 是各时点指标数值；f 是间隔的时间。

【例 6-2】

某企业 2009 年 4 月 1 日在册职工 300 人，4 月 11 日新进职工 9 人，而到 4 月 16 日又离开 4 人，其余时间职工人数没有变化，则该企业 4 月份平均职工人数为：

$$\bar{a} = \frac{300 \times 10 + 309 \times 5 + 305 \times 15}{10 + 5 + 15} = 304(\text{人})$$

2）根据间断时点数列计算序时平均数

在间断时点数列中有间隔相等和间隔不等两种情况：

- 对间隔相等的间断时点数列求序时平均数。当掌握资料是间隔相等的期初或期末资料时，序时平均数的计算方法是：假设在两时点间的变动是均匀的。将期初数量加期末数量，除以 2，即得本身的序时平均数；然后，再将各段时期的序时平均数相加，除以时期数，则得整个时期的序时平均数。

➪【例 6-3】

某企业 2008 年第二季度商品库存额资料如表 6-3 所示。

表 6-3　某企业 2008 年第二季度商品库存额记录表　　单位：万元

日期	3 月末	4 月末	5 月末	6 月末
商品库存额	100	86	104	114

根据表 6-3 可计算各月和第二季度的平均商品库存额：

$$4\text{ 月平均库存额}=\frac{100+86}{2}=93\text{（万元）}$$

$$5\text{ 月平均库存额}=\frac{86+104}{2}=95\ (\text{万元})$$

$$6\text{ 月平均库存额}=\frac{104+114}{2}=109\text{（万元）}$$

$$\text{第二季度平均库存额}=\frac{93+95+109}{3}=99\text{（万元）}$$

将上述计算第二季度平均库存额的两个步骤合并为：

$$\text{第二季度平均库存额}=\frac{\frac{100+86}{2}+\frac{86+104}{2}+\frac{104+114}{2}}{3}$$

$$=\frac{93+95+109}{3}=99\text{（万元）}$$

将上面计算过程概括为一般公式：

$$\bar{a}=\frac{\frac{a_1+a_2}{2}+\frac{a_2+a_3}{2}+\cdots+\frac{a_{n-1}+a_n}{2}}{n-1}$$

$$=\frac{\frac{a_1}{2}+a_2+a_3+\cdots a_{n-1}+\frac{a_n}{2}}{n-1}\qquad(6\text{-}3)$$

式中，$\bar{a}$ 是序时平均数；a 是各时点指标数值；n 是时点项数。

这种计算方法也称为“首末折半法”。

- 对间隔不等的间断时点数列求序时平均数。在时点数列中，如果相邻时点间隔不等时，就须首末折半后用相应的时点间隔数加权计算。其计算公式为：

$$\bar{a}=\frac{\frac{a_1+a_2}{2}f_1+\frac{a_2+a_3}{2}f_2+\cdots+\frac{a_{n-1}+a_n}{2}f_{n-1}}{\sum_{i=1}^{n-1}f_i} \quad (6\text{-}4)$$

式中，$\bar{a}$ 是序时平均数；a 是各时点指标数值；f 是各时点间隔的距离。

➪【例 6-4】

某地区 2008 年城乡居民储蓄存款余额资料如表 6-4 所示。

表 6-4　某地区 2008 年城乡居民储蓄存款余额记录表　　单位：万元

日　期	1 月 1 日	3 月 1 日	7 月 1 日	8 月 1 日	10 月 1 日	12 月 31 日
存款余额	38	42	54	56	60	71

根据上表资料计算的该地区 2008 年城乡居民平均储蓄存款余额为：

$$\bar{a}=\frac{\frac{38+42}{2}\times2+\frac{42+54}{2}\times4+\frac{54+56}{2}\times1+\frac{56+60}{2}\times2+\frac{60+71}{2}\times3}{2+4+1+2+3}=53.29 \text{（万元）}$$

根据间断时点数列计算序时平均数，是假定研究现象在相邻两个时点之间的变动是均匀的，实际上各种现象的变动一般是不均匀的。所以，其计算结果具有一定程度的假定性。因此，要使计算结果尽量反映实际情况，间断时点数列的间隔不宜过长，以缩小计算误差。

2．由相对数时间数列计算序时平均数

相对数时间数列是派生数列，即其中各项指标数值都是由两个绝对数对比计算出来的。按照数列的性质，要求利用其相应的两个绝对数时间数列，分别计算分子数列的序时平均数和分母数列的序时平均数，而后加以对比，即可求得。其计算公式为：

$$\bar{c}=\frac{\bar{a}}{\bar{b}} \quad (6\text{-}5)$$

式中，$\bar{c}$ 是相对数时间数列的序时平均数；$\bar{a}$ 是分子数列的序时平均数；$\bar{b}$ 是分母数列的序时平均数。

具体计算时又分以下几种情况。

（1）由两个时期数列对比而成的相对数时间数列求序时平均数

若构成相对数时间数列的分子数列和分母数列为时期数列，则分别用时期数列的公式，计算其序时平均数。其计算公式为：

$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{\frac{\sum a}{n}}{\frac{\sum b}{n}}=\frac{\sum a}{\sum b} \tag{6-6}$$

【例 6-5】

某企业 2008 年第一季度各月的计划产量及计划完成程度如表 6-5 所示。计算该企业第一季度平均月计划完成程度。

表 6-5　某企业 2008 年第一季度各月的计划产量及完成程度

月份	1	2	3
计划完成程度（%）	90	110	120
计划产量（件）	600	700	720

在此，必须注意，不能直接用各月产量计划完成程度相加，除以 3 求得，而必须先求出各月的实际产量（实际产量=计划产量×计划完成程度）。

1 月实际产量=600×90%=540（件）

2 月实际产量=700×110%=770（件）

3 月实际产量=720×120%=864（件）

然后计算第一季度平均月实际产量：

$$\bar{a}=\frac{\sum a}{n}=\frac{540+770+864}{3}=\frac{2\,174}{3}=724.7\text{（件）}$$

用同样的方法，可计算第一季度平均月计划产量：

$$\bar{b}=\frac{\sum b}{n}=\frac{600+700+720}{3}=\frac{2\,020}{3}673.3\text{（件）}$$

将 $\bar{a}$ 与 $\bar{b}$ 对比，则得第一季度平均月计划完成程度：

$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{724.7}{673.3}=107.6\%$$

（2）由两个时点数列对比而成的相对数时间数列求序时平均数

若构成相对数时间数列的分子数列与分母数列都是间隔相等的时点数列，用式（6-3）先分别求出 a 数列的 $\bar{a}$ 和 b 数列的 $\bar{b}$，然后求 $\bar{c}=\bar{a}/\bar{b}$；若两个都是间隔不等的时点数列，则用式（6-4）先分别求出 $\bar{a}$ 和 $\bar{b}$，然后求 $\bar{c}=\bar{a}/\bar{b}$。

➪【例 6-6】

某企业 2008 年第一季度工人人数与职工总数资料如表 6-6 所示。计算该企业第一季度工人占全部职工总数的平均比重。

表 6-6　某企业 2008 年第一季度工人人数与职工总数　　单位：人

日/月	1/1	1/2	1/3	1/4
工人人数	1 400	1 650	1 794	1 760
职工总数	2 000	2 200	2 400	2 200

利用式（6-3），则可计算第一季度工人占全部职工总数的平均比重：

$$\bar{c}=\frac{\dfrac{\frac{1\,400}{2}+1\,650+1\,794+\frac{1\,760}{2}}{4-1}}{\dfrac{\frac{2\,000}{2}+2\,200+2\,400+\frac{2\,200}{2}}{4-1}}=\frac{1\,675}{2\,233}=75\%$$

（3）一个时期数列和一个时点数列对比而成的相对数时间数列求序时平均数

此时，计算序时平均数的基本公式仍是 $\bar{c}=\bar{a}/\bar{b}$，至于 $\bar{a}$ 和 $\bar{b}$ 的具体计算方法则根据分子数列和分母数列的性质、类别而定。

➪【例 6-7】

某商业企业 2008 年第一季度商品销售额与库存额的资料如表 6-7 所示，计算该商业企业第一季度平均商品流转次数。

表 6-7　某商业企业 2008 年第一季度商品销售额与库存额

月份	1	2	3	4
商品销售额（万元）	120	143	289	—
月初商品库存额（万元）	50	70	60	110

第一季度平均月商品销售额可按下式计算：

$$\bar{a}=\frac{120+143+289}{3}=184\text{（万元）}$$

第一季度平均月商品库存额可按下式计算：

$$\bar{b}=\frac{\frac{50}{2}+70+60+\frac{110}{2}}{4-1}=\frac{210}{3}=70\text{（万元）}$$

第一季度平均商品流转次数：

$$\bar{c}=\frac{\bar{a}}{\bar{b}}=\frac{184}{70}=2.63\text{（次）}$$

3．由平均数时间数列计算序时平均数

平均数时间数列也是由两个绝对数时间数列派生的，因此要计算平均数时间数列的序时平均数，也须先求出分子、分母两个数列的序时平均数，然后对比计算。其计算公式仍为式（6-5）。

由于平均数时间数列所属性质的不同，可将其分为：由一般平均数组成的时间数列和由序时平均数形成的时间数列。在计算方法上它们是不同的。

（1）由一般平均数组成的时间数列计算序时平均数

其计算方法与由相对数时间数列计算序时平均数的方法一样，即先求出分子、分母数列的序时平均数，然后用式（6-5）求得。

【例 6-8】

某企业 2008 年第一季度工人人数和工业总产值资料如表 6-8 所示，计算该企业第一季度平均月劳动生产率。

表 6-8　某企业 2008 年第一季度工人人数和工业总产值

月份	1	2	3	4
工人人数（人）	1 850	2 050	1 950	2 150
工业总产值（万元）	250	272	271	—

第一季度平均月工业总产值为：

$$\bar{a}=\frac{250+272+271}{3}=264.33\text{（万元）}$$

第一季度平均月工人数为：

$$\bar{b}=\frac{\frac{1850}{2}+2\,050+1\,950+\frac{2150}{2}}{4-1}=\frac{6\,000}{3}=2\,000\text{（人）}$$

第一季度平均月劳动生产率：

$$\bar{c}=\frac{\bar{a}}{b}=\frac{264.33}{2\,000}=0.132\,167\text{（万元）}=1\,321.67\text{（元）}$$

（2）由序时平均数组成的时间数列计算序时平均数

根据序时平均数时间数列计算序时平均数，如果数列中各个时期的间隔相等，可用简单算术平均法计算序时平均数；如果数列中各个时期的间隔不等，则以间隔长度为权数用加权算术平均法计算序时平均数。现分别举例说明如下。

⇨【例6-9】

某水泥厂2008年各季度平均月产量资料如表6-9所示。

表6-9　某水泥厂2008年各季度平均月产量　　单位：万吨

季度	1	2	3	4
平均月产量	230	270	280	260

上表中各季水泥产量时期相等，可用简单算术平均法计算全年平均月产量：

$$\text{全年平均月产量}=\frac{230+270+280+260}{4}=260\text{（万吨）}$$

⇨【例6-10】

2008年某旅游区旅客的月平均人数资料如表6-10所示。

表6-10　2008年某旅游区旅客的月平均人数表　　单位：万人次

月份	1	2～3	4～8	9～11	12
月平均人次	10	13	16	14	10

上表中旅游月平均人次的时期不等，应采用加权算术平均法计算序时平均数：

$$\text{全年平均每月人数}=\frac{10\times1+13\times2+16\times5+14\times3+10\times1}{1+2+5+3+1}$$

$$=\frac{168}{12}=14\text{（万人次）}$$

6.2.3　增长量和平均增长量

1．增长量

增长量是说明社会经济现象在一定时期内所增长的绝对数量，它是报告期水平与基期水平之差，反映报告期比基期增长的水平。其计算公式为：

$$\text{增长量}=\text{报告期水平}-\text{基期水平} \tag{6-7}$$

由于采用的基期不同，增长量可分为逐期增长量和累计增长量。逐期增长量是指报告期水平与前一期水平之差，它表明本期比上一期增长的绝对数量；累计增长量是指报告期水平与某一固定时期（基期）水平之差，它表明本期比某一固定时期增长的绝对数量，即说明在某一段较长时期内总的增长量。

这两个指标可用如下形式表示：

逐期增长量：　　$a_1-a_0, a_2-a_1, \cdots, a_n-a_{n-1}$

累计增长量：　　$a_1-a_0,\ a_2-a_0, \cdots, a_n-a_0$

逐期增长量与累计增长量的关系是：

- 逐期增长量之和等于累计增长量，即：

$$(a_1-a_0)+(a_2-a_1)+\cdots+(a_n-a_{n-1})=a_n-a_0$$

- 两个相邻累计增长量之差等于逐期增长量。即：

$$(a_i-a_0)-(a_{i-1}-a_0)=a_i-a_{i-1}$$

在实际工作中，常计算年距增长量指标，它是报告期水平与上年同期水平之差，用于消除季节变动的影响。计算年距增长量可以消除季节变动的影响，表明报告期水平较上年同期水平增加（或减少）的绝对数量。

2．平均增长量

平均增长量说明社会经济现象在一定时期内平均每期增长的数量，从广义来说，它也是一种序时平均数，即是逐期增长量时间数列的序时平均数，反映现象平均增长水平。其计算公式为：

$$\text{平均增长量}=\frac{\text{逐期增长量之和}}{\text{逐期增长量个数}}=\frac{\text{累计增长量}}{\text{时间数列项数}-1} \tag{6-8}$$

➪【例 6-11】

2000—2004 年我国工业增加值资料如表 6-11 所示，现具体计算增长量和平均增长量指标。

表 6-11　2000—2004 年我国工业增加值　　单位：万元

年份		2000	2001	2002	2003	2004
工业增加值		39 570	42 607	45 935	53 612	62 815
增长量	逐期	—	3 037	3 328	7 677	9 203
	累计	—	3 037	6 365	14 042	23 245

$$\text{工业增加值年平均增长量}=\frac{3\,037+3\,328+7\,677+9\,203}{4}$$

$$=\frac{23\,245}{4}=5\,811.25\text{（万元）}$$

或

$$=\frac{23\,245}{5-1}=5\,811.25\text{（万元）}$$

6.3 时间数列的速度指标

根据时间数列可以计算的速度指标主要有：发展速度、增长速度、平均发展速度和平均增长速度。

6.3.1 发展速度和增长速度

1．发展速度

发展速度是两个时期发展水平对比而得的一种相对数，它表明现象发展的程度，说明报告期水平是基期水平的百分之几（或若干倍），通常用百分数表示。用公式表示为：

$$\text{发展速度}=\frac{\text{报告期水平}}{\text{基期水平}} \qquad (6\text{-}9)$$

由于采用的基期不同，发展速度可分为环比发展速度和定基发展速度。环比发展速度是以报告期水平与前一时期水平之比计算的发展速度，它用来说明报告期水平已经发展到了前一期水平的百分之几（或多少倍），表明这种现象逐期的发展程度。如果计算的单位时期为一年，这个指标也可叫做“年速度”。定基发展速度是指以报告期水平与某一固定时期水平之比计算的发展速度，它用来说明报告期水平已经发展到了固定时期水平的百分之几（或多少倍），表明这种现象在较长时期内总的发展程度，因此，有时也叫做“总速度”。这两种发展速度可用如下形式表示。

环比发展速度：　$\frac{a_1}{a_0},\frac{a_2}{a_1},\cdots,\frac{a_n}{a_{n-1}}$

定基发展速度：$\frac{a_1}{a_0},\frac{a_2}{a_0},\cdots,\frac{a_1}{a_0}$

环比发展速度和定基发展速度之间的关系表现如下。

定基发展速度等于环比发展速度的连乘积，即：

$$\frac{a_n}{a_0}=\frac{a_1}{a_0}\times\frac{a_2}{a_1}\times\frac{a_3}{a_2}\times\cdots\times\frac{a_n}{a_{n-1}}$$

两个相邻时期的定基发展速度之比，等于它们的环比发展速度，即：

$$\frac{a_n}{a_0}\div\frac{a_{n-1}}{a_0}=\frac{a_n}{a_{n-1}}$$

利用以上的关系，我们可以进行相互推算。

在实际工作中，还常要计算年距发展速度指标。它是报告期发展水平与上年同期发展水平之比。计算年距发展速度，也可以消除季节变动的影响，表明本期比上年同期的相对发展程度。

2．增长速度

增长速度是根据增长量与基期水平对比而求得的一种相对数。它表明现象在一定时期内增长的程度，说明报告期水平比基期水平增长了百分之几（或若干倍）。用公式表示为：

$$\text{增长速度}=\frac{\text{增长量}}{\text{基期水平}} \tag{6-10}$$

增长速度与发展速度具有密切关系，两者仅相差一个基数，所以将发展速度减 1（或 100%），就可求得增长速度。

$$\begin{aligned}\text{增长速度}&=\frac{\text{增长量}}{\text{基期水平}}=\frac{\text{报告期水平}-\text{基期水平}}{\text{基期水平}}\\&=\frac{\text{报告期水平}}{\text{基期水平}}-1=\text{发展速度}-1(\text{或}100\%)\end{aligned} \tag{6-11}$$

同样，由于采用的基期不同，增长速度也有环比增长速度和定基增长速度之分。

环比增长速度是报告期的逐期增长量与前一期发展水平之比，表明现象逐期的增长程度。其计算公式为：

$$\text{环比增长速度}=\frac{\text{逐期增长量}}{\text{前一期水平}} \tag{6-12}$$

或　　$=\text{环比发展速度}-1$

定基增长速度是报告期累计增长量与某一固定基期水平之比，它表明现象在较长时期内总的增长程度。其计算公式为：

$$\text{定期增长速度}=\frac{\text{累计增长量}}{\text{固定基期水平}} \qquad (6\text{-}13)$$

或 $$=\text{定基发展速度}-1$$

由此可见，发展速度大于1，则增长速度为正值，说明现象增长的程度时用“增加了”表示；反之，发展速度小于 1，则增长速度为负值，说明现象降低的程度时用“降低了”表示。

需要注意，上述两种增长速度不能直接相互推算，就是说，环比增长速度的连乘积不等于定基增长速度。如需推算，必须将增长速度加 1 或 100%化为发展速度，利用前述两种发展速度的数量关系式进行计算，然后再从计算结果中减去 1 或 100%，才得所求的增长速度。

【例 6-12】

我国 1996—2000 年水泥产量的发展速度和增长速度如表 6-12 所示。

表 6-12　我国 1996—2000 年水泥产量的发展速度和增长速度

年份		1996	1997	1998	1999	2000
水泥产量（万吨）		49 119	51 174	53 600	57 300	59 700
发展速度(%)	环比	—	104.2	104.7	106.9	104.2
	定基	100	104.2	109.1	116.7	121.5
增长速度(%)	环比	—	4.2	4.7	6.9	4.2
	定基	—	4.2	9.1	16.7	21.5

从表 6-13 中可看出，2000 年定基发展速度为 121.5%，而 1996～2000 年的环比发展速度的连乘积为：

$$104.2\%\times 104.7\%\times 106.9\%\times 104.2\%=121.5\%$$

正好等于 2000 年的定基发展速度。

另外，2000 年的环比发展速度为 104.2%，而 2000 年的定基发展速度与 1999 年的定基发展速度之比为：

$$\frac{121.5\%}{116.7\%}=104.2\%$$

也正好等于 2000 年的环比发展速度。

在实际工作中，我们也常计算年距增长速度，它是现象报告期的年距增长量与上年同期发展水平之比，或年距发展速度减 1 求得。用于说明年距增长量与上年同期发展水平对

比达到的相对增长程度。

6.3.2　平均发展速度和平均增长速度

为了观察社会经济现象在一个较长时期内逐期平均发展变化的程度和逐期平均增长变化的程度，就须计算平均发展速度和平均增长速度指标。平均速度指标是动态研究中很重要的两个分析指标。

1．平均发展速度

平均发展速度是各期环比发展速度的序时平均数。由于环比发展速度是根据同一现象在不同时间发展水平对比而得的动态相对数，因此它不能应用上述所讲的计算序时平均数的方法来计算。在实际工作中，计算平均发展速度的方法主要有两种，即几何平均法和方程法。两种方法数理依据不同，具体计算和应用场合也不一样。现分述之。

（1）几何平均法

计算平均发展速度时，因为总速度不等于各期环比发展速度的算术总和，而等于各期环比发展速度的连乘积，所以不能应用算术平均法，而要应用几何平均法来计算。在实践中，如果用水平法制定长期计划，则要求用几何平均法计算其平均发展速度。按此平均发展速度发展，可以保证在最后一年达到规定的 a_n 水平，所以几何平均法也称为“水平法”。即从最初水平 a_0 出发，以平均发展速度 $\bar{x}$ 代替各环比发展速度 x_1，x_2，x_3，…，x_n，经过 n 期发展，正好达到最末水平 a_n，用公式表示如下：

$$a_0\ x_1\ x_2\ x_3\cdots x_n = a_n$$

$$a_0 \underbrace{\bar{x}\ \bar{x}\ \bar{x}\cdots\bar{x}}_{n\text{个}} = a_n$$

$$\bar{x}^n = \frac{a_n}{a_0}$$

因此，平均发展速度 $\bar{x}$ 计算公式为：

$$\bar{x} = \sqrt[n]{\frac{a_n}{a_0}} \tag{6-14}$$

因为 a_n/a_0 为 n 期的定基发展速度，根据定基发展速度等于相应时期各环比发展速度连乘积的关系，所以计算平均发展速度也可以用下列公式：

$$\bar{x} = \sqrt[n]{\frac{a_n}{a_0}} = \sqrt[n]{x_1x_2x_3\cdots x_n} = \sqrt[n]{\prod x} \tag{6-15}$$

又因为 a_n/a_0 也是整个时期的总速度，所以平均发展速度还可以根据总速度计算，公式

如下：

$$\bar{x}=\sqrt[n]{\frac{a_n}{a_0^n}}=\sqrt[n]{R} \tag{6-16}$$

式中，$\bar{x}$ 是平均发展速度；x 是各期环比发展速度；$\prod$ 是连乘符号；R 是总速度；n 是环比发展速度的项数。

【例 6-13】

以表 6-12 我国 1996—2000 年水泥产量资料为例。

根据最初水平和最末水平计算平均发展速度：

$$\bar{x}=\sqrt[n]{\frac{a_n}{a_0}}=\sqrt[4]{\frac{59\,700}{49\,119}}=105\%$$

根据各环比发展速度计算平均发展速度：

$$\begin{aligned}\bar{x}&=\sqrt[n]{\prod x}\\&=\sqrt[4]{104.2\%\times104.7\%\times106.9\%\times104.2\%}\\&=105\%\end{aligned}$$

根据总速度计算平均发展速度：

$$\bar{x}=\sqrt[n]{R}=\sqrt[4]{121.5\%}=105\%$$

（2）方程法

在实践中，如果长期计划按累计法制定，则要求用方程法计算平均发展速度，按此平均发展速度发展，可以保证在计划内各期发展水平的累计达到计划规定的总数，所以方程法也称累计法。即从最初水平 a_0 出发，各期按平均发展速度 $\bar{x}$ 计算发展水平，则计算的各期发展水平累计总和，应与实际所具有的各期发展水平的累计总和相等。列出方程式，再求解便得出平均发展速度。

设 $\bar{x}$ 为平均发展速度，按平均发展速度计算的各期水平的假定值如下。

第 1 期：$a_1=a_0\bar{x}$

第 2 期：$a_2=a_0\bar{x}\,\bar{x}=a_0\bar{x}^2$

第 3 期：$a_3=a_0\bar{x}^2\bar{x}=a_0\bar{x}^3$

⋮

第 n 期：$a_n=a_0\bar{x}^{n-1}\bar{x}=a_0\bar{x}^n$

故各期假定水平之和为：

$$a_0\bar{x}+a_0\bar{x}^2+a_0\bar{x}^3+\cdots+a_0\bar{x}^n=a_0(\bar{x}+\bar{x}^2+\bar{x}^3+\cdots\bar{x}^n)$$

因各期实际水平之和为：

$$a_1+a_2+a_3+\cdots+a_n=\sum_{i=1}^{n}a_i$$

两者相等，则可列如下方程式：

$$a_0(\bar{x}^n+\bar{x}^{n-1}+\cdots+\bar{x}^2+\bar{x})=\sum_{i=1}^{n}a_i$$

即：

$$\bar{x}^n+\bar{x}^{n-1}+\cdots+\bar{x}^2+\bar{x}=\frac{\sum_{i=1}^{n}a_i}{a_0}$$

解此方程所得的正根就是要计算的平均发展速度。但是要解这个高次方程是比较复杂的，实际工作中都是根据事先编好的《平均增长速度查对表》来计算的。

下面，我们举例说明如何应用查表法来求平均发展速度和平均增长速度。

2．平均增长速度

平均增长速度是各期环比增长速度的序时平均数，它表明现象在一定时期内逐期平均增长变化的程度。根据增长速度与发展速度之间的运算关系，要计算平均增长速度，首先要计算出平均发展速度指标，然后将其减 1（或 100%）求得。即：

$$\text{平均增长速度}=\text{平均发展速度}-1(\text{或}100\%)$$

平均发展速度大于“1”，平均增长速度就为正值，表示某种现象在一个较长时期内逐期平均递增的程度，这个指标也叫做“平均递增速度”或“平均递增率”；反之，平均发展速度小于“1”，平均增长速度为负值，表示某种现象在一个较长时期内逐期平均递减程度，这个指标也可叫做“平均递减速度”或“平均递减率”。

➪【例 6-14】

某地区 2002—2007 年住宅建设投资额如表 6-13 所示。

表 6-13　某地区 2002—2007 年住宅建设投资额　　单位：亿元

年份	2002	2003	2004	2005	2006	2007
住宅建设投资额	10.0	11.0	12.0	13.5	14.0	14.5

第一步，计算递增或递减速度：

$$\frac{a_1+a_2+a_3+a_4+a_5}{n}=\frac{11.0+12.0+13.5+14.0+14.5}{5}=\frac{65}{5}=13$$

若（$a_1+a_2+a_3+a_4+a_5$）/$n>a_0$，则为递增速度，应在累计法查对表的增长速度部分查找。

第二步，计算总发展速度：

$$\frac{a_1+a_2+a_3+a_4+a_5}{a_0}=\frac{65}{10}=6.5=650\%$$

第三步，查表。见表6-14，在累计法查对表中的n=5的栏内，找到接近650%数值介于648.55%～650.44%之间，再查到该数所在行左边第一栏内的百分比介于8.8%～8.9%之间，即平均每年增长速度也即该地区“九五”期间住宅建设投资平均每年发展速度在108.8%～108.9%之间。运用内插法，还可得到更具体的数值。计算过程如下：

$$\frac{650\%-648.55\%}{650.44\%-648.55\%}=\frac{d}{8.9\%-8.8\%}$$

$$d=\frac{1.45\%\times 0.1\%}{1.89\%}=0.077\%$$

5年的平均增长速度为8.8% + 0.077% =8.877%；平均发展速度为108.877%。

表6-14　累计法查对表（间隔期：1～5年）

平均每年增长速度（%）	各年发展总和为基期的百分比（%）				
	1年	2年	3年	4年	5年
⋮			⋮		
8.8	108.80	227.17	355.96	496.08	648.53
8.9	108.90	227.49	356.63	497.26	650.41
⋮			⋮		
10.9	110.90	233.89	370.29	521.54	689.32
11.0	111.00	234.21	370.97	522.77	691.27
⋮			⋮		

3．计算和运用平均发展速度时应注意的问题

（1）根据统计研究目的选择计算方法

前述计算平均发展速度有几何平均法（水平法）和方程法（累计法），这两种方法在具

体运用上各有其特点和局限性。当目的在于考察最末一年发展水平而不关心各期水平总和时，可采用水平法。当目的在于考察各期发展水平总和而不关心最末一年水平时，可采用累计法。这样可以扬长避短，发挥两种计算方法的作用。

（2）要注意社会经济现象的特点

- 当现象随着时间的发展比较稳定地逐年上升或逐年下降时，一般采用水平法计算平均发展速度。但要注意，如果编制的时间数列中，最初水平和最末水平受特殊因素的影响而出现过高或过低的情况，则不可计算平均发展速度。
- 当现象的发展不是有规律地逐年上升或下降，而是经常表示为升降交替，一般采用累计法计算平均发展速度。但要注意，如果资料中间有几年环比速度增长得特别快，而有几年又降低得较多，出现了显著的悬殊和不同的发展方向，不易于计算出平均发展速度。因为用这样的资料计算的平均发展速度会降低这一指标的意义，不能确切说明实际情况。

（3）应采取分段平均速度来补充说明总平均速度

这在分析较长历史时期资料时尤为重要。因为仅根据一个总的平均速度指标只能笼统概括地反映其在很长时期内逐年平均发展或增长的程度，对深入了解这种现象的发展过程和变化情况往往是不够的。例如，要分析新中国成立以来粮食产量的平均发展速度和平均增长速度时，就有必要分别以国民经济恢复时期、各个五年计划时期和各个特定时期（如某几年受自然灾害的影响，产量逐年下降）等分段计算其平均速度加以补充说明。

（4）平均速度指标要与其他指标结合应用

- 要与发展水平、增长量、环比速度、定基速度等各项基本指标结合应用，起到分析、研究和补充说明的作用，以便对现象有比较确切和完整的认识。
- 在经济分析中，要与其他有关经济现象的平均速度指标结合运用。例如，工农业生产的平均速度、基本建设投资额与新增固定资产的平均速度、商品销售额与利润额的平均速度等，都可结合进行比较研究，以便深入了解有关现象在各个研究时期中每年平均发展和增长的程度等，为研究国民经济各种具有密切联系的现象的发展动态提供数据。

6.3.3　增长 1%的绝对值

增长速度反映了经济现象增长的相对程度，而增长量反映了经济现象增长的绝对数量。在对经济现象的增长情况进行统计分析时，往往将绝对数与相对数结合起来分析，即将增长 1%绝对值与速度指标结合起来进行统计分析。增长 1%绝对值是指逐期增长量与环比增长速度的比值。速度指标数值的大小与基期水平的高低密切相关，通常基期水平越高，发展速度增长 1%的绝对值就越大。增长 1%绝对值的计算公式为：

$$\text{增长1\%的绝对值}=\frac{\text{逐期增长量}}{\text{环比增长速度}}=\frac{a_n-a_{n-1}}{\frac{a_n-a_{n-1}}{a_{n-1}}\times 100}=\frac{a_{n-1}}{100}$$

6.4 现象变动的趋势分析

6.4.1 现象变动趋势分析的意义

时间数列各项发展水平的变化，是由许多复杂因素共同作用的结果。影响因素归纳起来大体有四类。

1．长期趋势

长期趋势指现象在一段较长的时间内，由于普遍的、持续的、决定性的基本因素的作用，使发展水平沿着一个方向，逐渐向上或向下变动的趋势。例如，粮食生产由于种植方法的不断改良、日益发达的农田水利设施等根本因素的影响，从较长时期来看，总趋势是持续增加、向上发展的。认识和掌握事物的长期趋势，可以把握事物发展变化的基本特点。

2．季节变动

季节变动指现象受季节影响而发生的变动。其变动的特点是，在一年或更短的时间内随着时序的更换，使现象呈周期重复的变化。引起季节变动的原因既有自然因素，也有人为因素，如气候条件、节假日以及风俗习惯等。季节变动的影响有以一年为周期的，也有以月、周、日为变动周期的。认识和掌握季节变动，对于近期行动决策有重要的作用。

3．循环变动

循环变动指现象发生周期比较长的涨落起伏的变动。通常所指的循环变动是经济发展荣衰交替的变动。它与春夏交替相继不息的天时循环变动有明显的不同，也不同于朝单一方向持续发展的长期趋势。循环变动可能由于不同的原因，使得变动的周期长短不同，常在一年以上，甚至七八年、十多年。各期始末难定为何年何月，上下波动程度也不相同。

4．不规则变动

不规则变动指现象除了受以上各种变动的影响以外，还受临时的、偶然因素或不明原因而引起的非周期性、非趋势性的随机变动。不规则变动是无法预知的。

现象变动趋势分析就是要把时间数列受各类因素的影响状况分别测定出来，弄清楚研究对象发展变化的原因及其规律，为预测未来和决策提供依据。

一般来说，时间数列变动包含了上述四类因素的影响，因而时间数列的成分结构包括这四种变动形式。但就实际情况看，季节变动和循环变动在某些场合并不存在。比如，按年排列的动态数列就不体现季节变动。我国工农业生产发展趋势一般不存在循环变动。因



此，在实际工作中，要对研究现象进行具体分析，实际包含什么因素就测定什么因素。在这里我们仅介绍通常使用的两种测定方法，即长期趋势和季节变动的测定。

6.4.2　长期趋势的测定

长期趋势的测定就是用一定的方法对时间数列进行修匀，使修匀后的数列排除季节变动、循环变动和无规则变动等因素的影响，显示出现象变动的基本趋势，作为预测的依据。测定长期趋势的方法主要有时距扩大法、移动平均法和数学模型法。数学模型又有线性模型和非线性模型之分。以下分别加以说明。

1．时距扩大法

这是对长期的时间数列资料进行统计修匀的一种简便方法。它是把原有动态数列中各时期资料加以合并，扩大每段计算所包括的时间，得出较长时距的新时间数列，以消除由于时距较短受偶然因素影响所引起的波动，清楚地显示现象变动的趋势和方向。

时距扩大法把较小的时间跨度转化为较大时间跨度，如昼夜转化为星期或旬，由旬转化为月，月转成季或年，由一年转化为许多年。如果数列水平波动有一定的周期性，扩大的时距应注意与各次摆动的周期相同；如果时间数列看不出有什么周期性，那么就要逐步扩大时距，直到趋势的方向变得足够清晰为止。

时距扩大修匀可以用扩大时距后的绝对数表示，也可以用扩大时距后的平均数表示。前者只适用于时期数列，后者可以用于时期数列和时点数列。

⇨【例 6-15】

表 6-15 是我国 1961—1995 年粮食产量资料，以它来说明时距扩大法的应用。

表 6-15　我国 1961—1995 年粮食产量　　单位：万吨

年份	产量	年份	产量	年份	产量
1961	14 750	1973	26 494	1985	37 911
1962	16 000	1974	27 527	1986	39 151
1963	17 000	1975	28 452	1987	40 298
1964	18 750	1976	28 631	1988	39 408
1965	19 453	1977	28 273	1989	40 755
1966	21 400	1978	30 477	1990	43 498
1967	21 782	1979	33 212	1991	43 524
1968	20 906	1980	32 056	1992	44 258
1969	21 097	1981	32 502	1993	45 644

续表

年份	产量	年份	产量	年份	产量
1970	23 996	1982	35 450	1994	44 450
1971	25 014	1983	38 728	1995	45 600
1972	24 048	1984	40 731		

从表中看出，35年来我国粮食产量呈不断增长的趋势，但中间有过几次波动。我们把时距扩大为5年，可消除短时间受偶然因素影响所带来的波动，如表6-16所示。

表6-16　时距扩大为5年后的粮食产量资料　　单位：万吨

年份	总产量	平均年产量
1961—1965	85 935	17 190.6
1966—1970	109 181	21 836.6
1971—1975	131 535	26 306.0
1976—1980	152 649	30 529.8
1981—1985	185 320	37 064.0
1986—1990	203 110	40 622.0
1991—1995	223 476	44 695.2

表6-16中把时距扩大为5年，把中间个别年份波动修匀了，形成35年来完全上升的总趋势。通过移动平均数对数列修匀，可以更深刻地描述现象发展的基本趋势。

2．移动平均法

移动平均法是采用逐期递推移动的方法计算一系列扩大时距的序时平均数，并以这一系列移动平均数作为对应时期的趋势值。通过移动平均数对数列修匀，可以消除原数列中的短期波动，更深刻地描述现象发展的长期趋势。

移动平均法所采用的扩大时距，也应由时间数列的具体特点决定。同时与时距扩大法一样，要注意数列水平波动的周期性。一般要求扩大的时距与周期变动的时距相吻合，或为它的整倍数，比如，对于具有季度水平资料的时期数列，经受每年季节性的涨落，主要必须消除季节变动因素，以运用4项或8项移动平均为宜。在以年为单位的数据所形成的时间数列中，不存在季节变动因素，因此要消除的是循环变动和不规则变动因素。我们可以借助于对时间数列水平的观察，循环周期大体几年，就相应采用几年移动平均。若数列水平呈无规则的波动，也是采取逐步扩大时距的办法，直到所求的移动平均数能把现象变动趋势表现出来。

移动平均法的具体做法是从时间数列第一项数值开始，按一定项数求序时平均数，逐

项移动，得出一个由移动平均数构成的新的时间数列，这个派生数列把受某些偶然因素影响所出现的波动修匀了，使整个数列的总趋势更加明显。移动平均法根据资料的特点及研究的具体任务，可能进行 3 项、4 项、5 项乃至更多项移动平均。数列项数为奇数项时，移动平均所得的数值放在中间一项的位置上，一次移动即可得到趋势值；偶数项移动平均所得的数值放在中间两项位置中间，并需要将第一次得到的趋势值进行二次移正平均，才能得到新的趋势值并组成新的时间数列。被移动平均的项数越多，对原数列修匀的作用就越大，但得到的新时间数列的项数却越少。

【例 6-16】

表 6-17 是我国 1995—2004 年啤酒产量资料。

根据资料进行 3 项和 4 项移动平均。

3 项移动平均：第一个平均数为（58 478.1+ 67 884.6 + 74 462.6）÷3 = 66 941.767，对正第 2 年的原值。以此类推移动平均，得出 3 年移动平均数列 8 项。

表 6-17　我国 1995—2004 年啤酒产量　　单位：万吨

年份	序号	产量	移动平均趋势值		
			3 项移动平均	4 项移动平均	5 项移动平均
1995	1	58 478.1			
1996	2	67 884.6	66 941.767		
				69 792.625	
1997	3	74 462.6	73 564.133		72 741.3
				75 689.975	
1998	4	78 345.2	78 291.767		78 384.7
				81 079.375	
1999	5	82 067.5	83 284.967		83 935.9
				86 792.425	
2000	6	89 442.2	89 608.167		90 098.1
				93 403.775	
2001	7	97 314.8	97 182.533		97 732.1
				102 060.4	
2002	8	104 709.6	106 266.467		107 944.5
				113 828.6	

续表

年份	序号	产量	移动平均趋势值		
			3 项移动平均	4 项移动平均	5 项移动平均
2003	9	116 694.0	119 333.2		
2004	10	136 515.0			

4 项移动平均：第一个平均数为(58 478.1+ 67 884.6 + 74 462.6+78 345.2) ÷ 4=69 792.625 万吨，对正第 2 ~ 3 项的中间。以此类推，得出 4 项移动平均数列。每个指标值位置与原数列相比都错半期，无法直接比较，因此还需进行一次移正平均。即再进行一次两项移动平均，这样各平均数都对准各期，形成新的 4 项平均移正平均数列。

从表 6-17 看出，移动平均的结果使短期的偶然因素引起的波动被削弱，整个时间数列被修匀得更加平滑，波动趋于平稳。

按移动平均法对时间数列修匀后，趋势值的个数比原数列实际水平的个数减少了。上面举例中，按 4 项和 5 项移动，首尾都有两个时期得不到趋势值。这无疑在一定程度上减少了研究最初和最末发展阶段显示趋势特点的可能性。所以移动平均法虽然能够看到长期趋势变动的特点，但是，移动平均数不能对趋势进行分析修匀，即无法对现象的发展趋势做出预测。

3．数学模型法

这是对时间数列进行分析修匀的常用方法，它用适当的数学模型对时间数列配合一个方程式，据以计算各期的趋势值。测定长期趋势广泛使用这种方法。下面分别介绍直线趋势和非直线趋势的测定。

（1）直线趋势的测定

如果时间数列逐期增长量相对稳定，即现象发展水平按相当固定的绝对速度变化时，则采用直线（线性函数）作为趋势线，来描述趋势变化，预测前景。

如以时间因素作为自变量（t），把数列水平作为因变量（y），配合的直线趋势方程为：

$$y_c = a + bt \qquad (6\text{-}17)$$

式中，y_c是时间数列的趋势值；t 是时间的序号。参数 a，b 的求法用最小平方法。

最小平方法是测定长期趋势最普遍使用的方法。它的原理是：数列实际值与数列的趋势值的离差平方和达到一个最小值，即 $\sum(y-y_c)^2$ =最小值。符合这个条件的只有一条线，所以这条线又称原数列的最适线，它使趋势线同原数列最佳配合。这条线也同时满足离差

之和为零的要求。

根据数学运算，上式可导出下列两个 a，b 求解标准方程：

$$\begin{cases}\sum y = na + b\sum t \\ \sum ty = a\sum t + b\sum t^2\end{cases}$$

解得：

$$a = \frac{\sum y}{n} - \frac{b\sum t}{n} = \bar{y} - b\bar{t}, b = \frac{\sum ty - \frac{1}{n}(\sum t)(\sum y)}{\sum t^2 - \frac{1}{n}(\sum t)^2} \qquad (6\text{-}18)$$

➪【例 6-17】

我国国内生产总值资料如表 6-18 所示，用最小平方法求国内生产总值的直线趋势方程。这里以 1，2，3，…，n 代表年份顺序。

表 6-18　最小平方法计算表　　单位：亿元

年份	时间代码 t	国内生产总值 y	t^2	ty	y_c=48 602+7 635.9t
1995	1	58 478.1	1	58 478.1	56 237.9
1996	2	67 884.6	4	135 769.2	63 873.8
1997	3	74 462.6	9	223 387.8	71 509.7
1998	4	78 345.2	16	313 380.8	79 145.6
1999	5	82 067.5	25	410 337.5	86 781.5
2000	6	89 442.2	36	536 653.2	94 417.4
2001	7	97 314.8	49	681 203.6	102 053.3
2002	8	104 790.6	64	838 324.8	109 689.2
2003	9	116 694.0	81	1 050 246.0	117 325.1
2004	10	136 515.0	100	1 365 150.0	124 961.0
合计	55	905 994.6	385	5 612 931.0	905 994.5

根据以上联立方程，求解得：

$$b = \frac{5\,612\,931 - \frac{1}{10} \times 55 \times 905\,994.6}{385 - \frac{1}{10} \times 55^2} = 7\,635.9$$

$$a=\frac{905\,994.6}{10}-\frac{7\,635.9\times 55}{10}=48\,602$$

代入直线方程：

$$y_c=a+bt=48\,602+7\,635.9t$$

该直线趋势方程表明，当 $t=0$ 时，趋势值为 48 602，即该直线的起点值；每增加一年，国内生产总值平均增长 76 35.9 亿元。将各年代码依次代入方程，可以得出趋势值数列（见表 6-18），并可进行外推预测。

国内生产总值直线趋势图如图 6-1 所示。

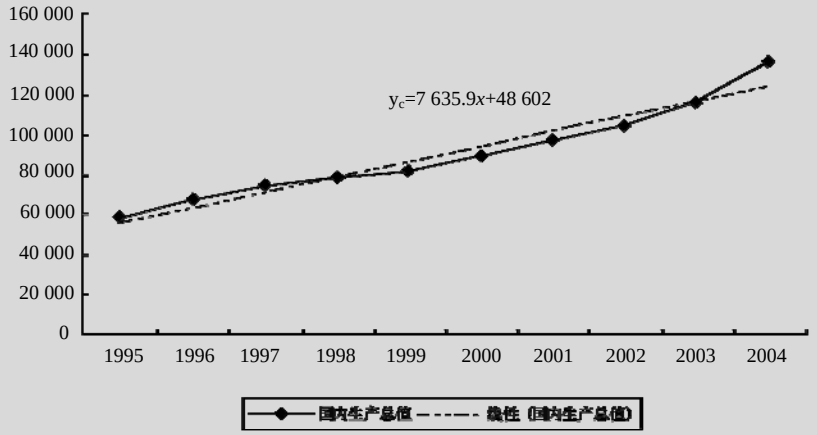


图 6-1　国内生产总值直线趋势图（最小平方法）

上例中，$t=0$ 时，表明直线趋势方程的原点在 1994 年。计算时可以取时间数列中任何一个时期为原点，如果开始的时间读数（$t=0$）采取中间的时期或时点，参数 a，b 计算就很简单。通常，当时间数列为奇数项时，可用中间项那一年为原点，比如，某数列有 9 项水平，时间跨度从 1990—1998 年，则 t 值分别为：

年份	1990	1991	1992	1993	1994	1995	1996	1997	1998
t	−4	−3	−2	−1	0	1	2	3	4

当时间数列为偶数项时，则可用两个中间项的中点为原点，这时就不是以年为测定的时间单位，而是以半年为测定的时间单位。比如，某数列有 8 项水平，时间跨度从 1991—1998 年，则 t 值做如下确定：

年份	1991	1992	1993	1994	1995	1996	1997	1998
t	−7	−5	−3	−1	1	3	5	7



当以上两种情况出现时，$\sum t=0$，标准方程简化为：

$$\begin{cases}\sum y = na \\ \sum ty = b\sum t^2\end{cases}$$

因此：

$$a=\frac{\sum y}{n}, b=\frac{\sum ty}{\sum t^2} \tag{6-19}$$

这样处理，计算过程可以大大简化。

（2）曲线趋势的测定

现实生活中，大量的现象是非线性发展的。因此，研究长期趋势变动的各种曲线类型是十分必要的。曲线类型很多，我们仅选定指数曲线来讨论非线性趋势的测定。

当时间数列大体上是按每期以相同的增长速度变化，即各期环比增长速度大体相同，则这种数列的基本趋势是指数曲线型，方程式为：

$$y_c = ab^t \tag{6-20}$$

进行指数曲线配合，必须先将指数曲线化为直线的形式。

对方程式 $y_c = ab^t$ 两边取对数，得：

$$\lg y = \lg a + t\lg b$$

令　　$y^t = \lg y,\ A = \lg a,\ B = \lg b$

则指数曲线的方程式可表示为直线形式：

$$y^t = A + Bt$$

于是，可以按直线配合的方法确定所需的指数曲线。

➪【例 6-18】

2002—2007 年某地的人口资料如表 6-19 所示，其年增长速度大体上一样，我们用最小平方法配合指数曲线方程。

表 6-19　2002—2007 年某地的人口

年份	人口 y（万人）	t	t^2	y^t=lgy	t lgy	y_c=84.547(1.011 3)^t
2002	85.50	1	1	1.932 0	1.932 0	85.50
2003	86.48	2	4	1.936 9	3.873 8	86.47
2004	87.46	3	9	1.941 8	5.825 4	87.45

续表

年份	人口 y（万人）	t	t^2	$y^t=\lg y$	$t\lg y$	$y_c=84.547(1.011\ 3)^t$
2005	88.47	4	16	1.946 8	7.787 2	88.43
2006	89.46	5	25	1.951 6	9.758 0	89.43
2007	90.44	6	36	1.956 4	11.738 4	90.44
合计	—	21	91	11.665 5	40.914 8	—

$$B=\frac{\sum t\lg y-\frac{1}{n}\sum t\sum \lg y}{\sum t^2-\frac{1}{n}(\sum t)^2}=\frac{40.914\,8-\frac{1}{6}\times 21\times 11.665\,5}{91-\frac{1}{6}\times(21)^2}=0.004\,89$$

$$b=1.011\ 3$$

由 $y^t=A+Bt$，得：

$$\frac{11.665\,5}{6}=A+0.004\,89\left(\frac{21}{6}\right)$$

$$A=1.927\ 1\ ,\ a=84.547$$

则配合的指数曲线方程如下：

$$y^t=1.927\ 1+0.004\ 89t$$

$$y_c=ab^t=84.547(1.011\ 3)^t$$

指数曲线方程中参数的意义是十分明确的。a 表示修匀数列的初始水平，b 表示 t 单位时间趋势值发展速度。比如，在上面的例子中，b=1.011 3 表示某地人口数的趋势值每年发展速度为 101.13%，或者说该地区人口平均每年以 1.13%速度递增。

6.4.3 季节变动的测定

1. 季节变动的含义

在日常经济生活中，我们经常会听到“销售旺季”、“旅游旺季”或“销售淡季”、“旅游淡季”之说，这是由于这些活动因季节的不同而发生着变化。所谓季节变动，是指客观现象由于受自然因素和生产或生活条件的影响在一年内随着季节的更换而引起的比较有规律的变动。

现在季节变动中的“季节”，不仅仅指一年中的四季，而是指任何一种周期性的变化。农业生产、旅游业、商品销售、交通运输等都有明显的季节变动规律。我们对现象的季节变动进行分析和研究，可以通过确定现象过去的季节变动规律来为当前的生产经营活动提供依据。此外，也可通过对季节变动的测定消除时间数列中的季节因素，以便分析其他构

成因素的影响。

季节变动是一种各年变化强度大体相同并且每年重复出现的有规律的变动，我们将其归纳为一种典型的季节模型。季节模型是由一套指数组成的，各个指数刻画了现象在一个年度内各月或季的典型数量特征。如果所分析的是月份数据，季节模型就由 12 个指数组成；如果是季度数据，季节模型就由 4 个指数组成。季节变动分析就是对一个时间序列计算出这种月（季）指数，即季节指数，也称季节比率，然后根据各季节指数与其平均数的偏差程度来测定季节变动的程度。下面将介绍常用的简单分析方法——按月（季）平均法。

2．按月（季）平均法

这种方法不考虑长期趋势影响，直接用原时间数列来计算。它是用各月（或季）的平均数作为该月（或季）的代表值，以消除随机影响，然后计算出各年总月（或总季）的平均数作为全年的代表值。二者相比，即为季节比率。其具体计算步骤如下。

- 根据各年按月（季）的时间数列资料计算出各年同月（季）的平均水平：

$$\bar{y}_i = \frac{\sum_{i=1}^{N} y_i}{N}$$（N 为年数）

- 计算各年所有月（季）的总平均水平：

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$（n 为月数或季数）

- 将各年同月（季）的平均水平与总平均水平进行对比，即得出季节比率：

$$S_i = \frac{\bar{y}_i}{\bar{y}} \tag{6-21}$$

【例 6-19】

某服装公司 2003—2007 年各月的销售额资料及计算的季节比率如表 6-20 所示。

表 6-20　某服装公司销售额及季节比率计算表

月份	各年销售额					5 年同月销售额合计（万元）	5 年同月销售额平均（万元）	季节比率（%）
	2003	2004	2005	2006	2007			
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
1	1.1	1.1	1.4	1.4	1.3	6.3	1.26	18.1

续表

月份	各年销售额					5年同月销售额合计（万元）	5年同月销售额平均（万元）	季节比率（%）
	2003	2004	2005	2006	2007			
	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
2	1.2	1.5	2.1	2.1	2.2	9.1	1.82	26.1
3	1.9	2.2	3.1	3.1	3.3	13.6	2.72	39.0
4	3.6	3.9	5.2	5.0	4.9	22.6	4.52	64.8
5	4.2	6.4	6.8	6.6	7.0	31.0	6.20	88.9
6	4.2	16.4	18.8	19.5	20.0	78.9	15.78	262.4
7	24.0	28.0	31.0	31.5	31.8	146.3	29.26	419.5
8	9.5	12.0	14.0	14.5	15.3	65.3	13.06	187.2
9	3.8	3.9	4.8	4.9	5.1	22.5	4.50	64.5
10	1.8	1.8	2.4	2.5	2.6	11.1	2.22	31.8
11	1.2	1.3	1.2	1.4	1.4	6.5	1.30	18.6
12	0.9	1.0	1.1	1.2	1.1	5.3	1.06	15.2
年总计	57.4	79.5	91.9	93.7	96.0	418.5	6.98	1 200.0

5年间月份的平均销售额为：

$$\bar{y}_i = \frac{\sum_{i=1}^{N} y_i}{N}$$

则1月份平均销售额为：

$$\bar{y} = \frac{1.1+1.1+1.4+1.4+1.3}{5} = 1.26\text{（万元）}$$

见表中第（7）栏数字。

5年间总平均月销售额为：

$$\bar{y} = \frac{\sum_{i=1}^{n} \bar{y}_i}{n}$$

$$= \frac{1.26+1.82+2.72+4.52+6.2+15.78+29.26+13.06+4.5+2.22+1.3+1.06}{12}$$

$$= 6.98(\text{万元})$$

季节比率为：

$$S_i = \frac{\overline{y}_i}{\overline{y}}$$

例如，1 月份的季节比率为：

$$S_1 = \frac{1.26}{6.98} = 18.1\%$$

2 月份的季节比率为：

$$S_2 = \frac{1.82}{6.98} = 26.1\%$$

见表中第（8）栏数字。

12 个月的季节比率之和应为 1 200%，4 个季度的季节比率之和应为 400%。如果不等，应当调整。其方法是：将 1 200（或 400）除以 12 个月（或 4 个季度）的季节比率之和，得到一个调整系数，然后，将此系数分别乘以原来的各个季节比率所得的数，即调整后的季节比率，它们之和为 1 200（或 400）。本例 12 个月的季节比率之和正好为 1 200，无须调整。

这样，由各月份季节比率组成的数列清楚地表明，该服装公司销售额的季节性变动趋势，自 1 月起逐月增长，7 月达到最高峰，8 月开始下降，到 12 月降到最低点。若以横轴表示月份，纵轴表示季节比率，绘成季节变动图，就更能明显地看出季节性变动趋势，如图 6-2 所示。

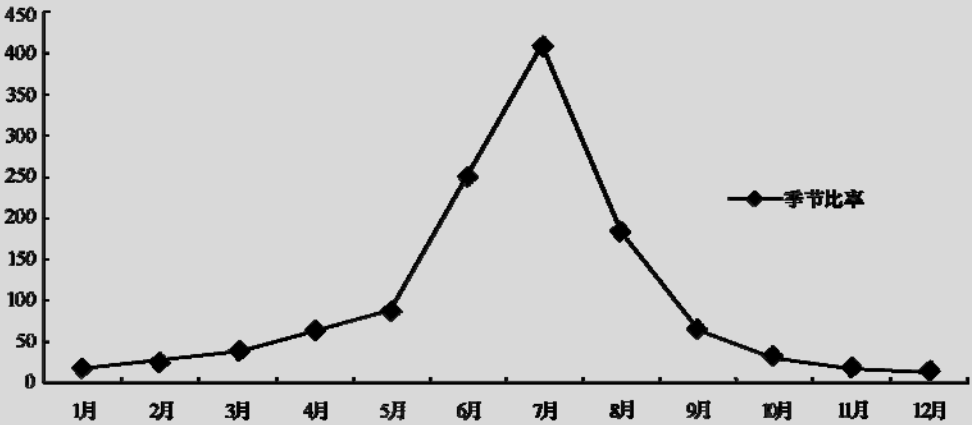


图 6-2　某服装公司销售额季节变动图

按月平均法计算简便，容易掌握。但季节比率的计算不够精确，因为它不考虑长期趋势的影响。在前后期月水平波动较大的资料中，后期各月水平比前期水平有较大提高，对平均数的影响大，从而影响了季节比率的准确性。我们可以用移动平均趋势剔除法来测定季节变动。在此，就不再介绍了。

应用季节变动的资料，可以进行某些外推预测。比如，时间数列没有明显的长期趋势，或允许不考虑长期趋势存在的情况下，可直接以按月（季）平均法计算的季节比率来调整各月（季）预测值。有两种方法：

第一，如果已测得下一年度全年预测值，则各月（季）预测值等于月（季）平均预测值乘以该月（季）的季节比率。

第二，如果已知下一年份几个月的实际水平，则以后各月（季）的预测值等于已知月（季）的实际水平乘以后各月（季）的季节比率与已知月（季）季节比率之比。

就上面关于服装销售的例子，假设已预测 2008 年全年销售额为 99.6 万元，平均每月销售额为 8.3 万元，则：

1 月份预测值为 8.3×17.6%=1.46（万元）

2 月份预测值为 8.3×25.5%=2.12（万元）

其余各月预测以此类推。

本章小结

1. 在社会经济统计中，时间数列可以用来反映社会经济发展变化的过程和特点，研究现象发展变化的趋势和规律，对未来状态进行科学的预测。

对时间数列可以从不同角度分类。从计算动态分析指标考虑，可按指标形式分为总量指标数列、相对指标数列和平均指标数列；从选择预测方法和确定预测模型考虑，可按指标数列形态分为随机性、平稳性、趋势型和季节型等数列。对时间数列性质和类型的识别，可以采用直观的曲线图，也可用自相关分析法。

2. 时间数列的分析指标常用的有平均发展水平、平均增长量、发展速度、增长速度、平均发展速度和平均增长速度等。这些指标不仅可以概括现象发展变化的过程和特点，而且是进行纵向、横向比较评价的依据。

3. 对时间数列进行分解，从中抽象出长期趋势、季节变动、循环变动和不规则变动，并加以测定，可以更具体地认识时间数列的特征和便于选择预测模型。

长期趋势量化测定的主要方法是时距扩大法、移动平均法以及最小二乘法；测定季节变动则有按月平均法和长期趋势剔除法。

4. 时间数列预测方法的主要特点，是把数列指标视为时间的函数或者是历史各期水平合乎规律变动的结果。它只是数列本身的外推预测，而不考虑影响数列变动的其他因素。时间数列预测常用的方法有趋势外推法、自回归预测法、指数平滑法等。现代时间数列分

析已经发展形成了许多更高级复杂的预测模型，必须借助于计算机技术进行操作。

复习思考题

一、概念题

时间数列　　序时平均数　　逐期增长量　　累计增长量　　环比发展速度　　定基发展速度　　平均发展速度　　增长率　　平均增长率

二、简答题

1. 什么是时间数列？编制时间数列应注意哪些基本要求？
2. 时期数列和时点数列有哪些不同的特点？
3. 进行动态水平和速度分析分别运用哪些指标？
4. 由时期数列和时点数列计算序时平均数有什么不同？当计算相对数或平均数时间数列的序时平均数时，应该怎样利用时期数列和时点数列计算上的特点来进行计算？
5. 环比发展速度和定基发展速度二者关系如何？环比增长速度和定基增长速度之间是否也存在相同的关系？
6. 什么是增长量？逐期增长量和累计增长量有何不同？二者关系如何？
7. 几何平均法与方程法计算平均发展速度有什么不同？哪些现象适用几何平均法？哪些现象适用方程法？举例说明。
8. 计算和运用平均发展速度应注意哪些问题？
9. 长期趋势测定的时距扩大法、移动平均法和数学模型法各自特点是什么？

三、练习题

1．判断题（把“√”或“×”填在题后的括号里）

（1）根据所采用指标的性质和表现形式的不同，动态数列可以分为绝对数动态数列、相对数动态数列、平均数动态数列。（　　）

（2）动态数列中的各项数值就是发展水平。（　　）

（3）国民生产总值翻两番就是增长一倍，即为基期的 200%。（　　）

（4）若环比增长速度每年相等，则其逐期增长量也年年相等。（　　）

（5）相对数动态数列、平均数动态数列中的各个指标数值是能相加的。（　　）

（6）若各期的逐期增长量△相等（△＞0），则各期的环比增长速度是逐年（期）下降的。（　　）

（7）各期的环比增长速度连乘积等于最末期的定期增长速度。（　　）

（8）某产品连续 5 年的成本平均发展速度为 98.9%，意味着某产品成本平均每年下降了 1.1。（　　）

2．单选题

（1）序时平均数与一般平均数的共同点是（　　）。

A. 两者都是反映同一总体的一般水平

B. 都是反映现象的一般水平

C. 两者均可消除现象波动的影响

D. 共同反映同质总体在不同时间上的一般水平

（2）下列属于时点数列的是（　　）。

A. 某企业各年总产值

B. 某企业年末拥有的设备数

C. 某企业年末生产工人数占全部职工人数的比重

D. 某企业月工资总额

（3）动态数列中的派生数列是（　　）。

A. 时期数列和时点数列

B. 绝对数动态数列和相对数动态数列

C. 绝对数动态数列和平均数动态数列

D. 相对数动态数列和平均数动态数列

（4）水平法平均发展速度的计算，是下列哪个指标的连乘积的 n 次方根。（　　）

A. 环比增长速度　　B. 环比发展速度

C. 定基发展速度　　D. 定基增长速度

（5）当时期数列分析的目的侧重于研究某现象在各时期发展水平的累计总和时，应采用（　　）。

A. 算术平均计算平均发展速度　　B. 调和平均计算平均发展速度

C. 累积法计算平均发展速度　　D. 水平法计算平均发展速度

（6）发展速度与增长速度的关系是（　　）。

A. 定基发展速度−1=环比增长速度

B. 环比发展速度−1=环比增长速度

C. 定期增长速度的连乘积=定基发展速度

D. 环比增长速度的连乘积=环比发展速度

（7）在时点数列中，称为“间隔”的是（　　）。

A. 最初水平与最末水平之间的距离

B. 最初水平与最末水平之差

C. 两个相邻指标在时间上的距离
D. 两个相邻指标数值之间的距离

（8）动态数列中的发展水平（　　）。
A. 只能是绝对数
B. 只能是相对数
C. 只能是平均数
D. 可以是绝对数，也可以是相对数或平均数

3．多选题

（1）下列指标构成的动态数列属于时期数列的是（　　）。
A. 全国每年大专院校毕业生人数
B. 某市全年新生婴儿数
C. 某商场各月末商品库存额
D. 某企业职工工资总额
E. 菜农场历年年末生猪存栏数

（2）时点数列的特点有（　　）。
A. 数列中各项指标数值相加之和有意义
B. 数列中各项指标数值相加之和没意义
C. 数列中各项指标数值的大小与其计算时间的长短存在着直接关系
D. 数列中各项指标数值的大小与其计算时间的长短不存在着直接关系
E. 数列中各项指标数值是间隔登记取得的

（3）把某企业 2005 年各月的总产值计划完成百分比，按月排列起来的数列称为（　　）。
A. 发展水平　　B. 变量数列　　C. 绝对数动态数列
D. 相对数时间数列　　E. 平均数时间数列

（4）用于分析现象发展速度的指标有（　　）。
A. 发展水平　　B. 平均发展水平　　C. 发展速度
D. 平均发展速度　　E. 增长量

（5）将不同时期的发展水平加以平均而得到的平均数称为（　　）。
A. 序时平均数　　B. 静态平均数　　C. 动态平均数
D. 一般平均数　　E. 平均发展水平

（6）定基增长速度等于（　　）。
A. 环比增长速度的连乘积　　B. 定期增长量除以固定水平
C. 定基发展速度减 1　　D. 环比增长量除以固定水平
E. 环比发展速度连乘积减去 100%

（7）水平法计算平均发展速度，根据所掌握的资料不同有如下计算公式（　　）。

A. $\sqrt[n]{\frac{a_n}{a_0}}$　　B. $\sqrt[n]{R}$　　C. $\frac{\sum x}{n}$

D. $\sqrt[n]{\prod_x}$　　E. $\frac{a_n}{na_0}$

4．计算题

（1）某自行车车库4月份有几次变动情况如表6-21所示。试计算该车库4月份平均库存自行车多少辆。

表6-21　某自行车车库4月份变动情况

4月1日	4月6日	4月8日	4月26日
320辆	调出70辆	进货120辆	调出80辆

（2）某企业1～7月份的资料如表6-22所示。

要求：

1）计算上半年月平均销售产值和平均职工人数。

2）计算上半年月平均全员劳动生产率、上半年的全员劳动生产率和上半年职工构成指标。

表6-22　某企业1—7月资料情况

月份	1	2	3	4	5	6	7
销售产值（万元）	84	102	90	92	91	94	--
期初职工人数（人）	100	98	100	103	105	105	110
其中：工人数（人）	68	60	68	72	75	75	80

（3）某水泥厂2003—2008年水泥产量如表6-23所示。

要求：

1）计算出表中动态分析指标各年的数值，并填入表内相应空格中。

2）计算平均增长量，并预测2009年该厂的水泥产量。

3）用水平法求2003—2008年水泥产量的平均发展速度和平均增长速度，并据之预测2009年该厂的水泥产量。

4）试用最小平均法配合趋势直线作图，并预测2009年的水泥产量。

表6-23　某水泥厂2003—2008年水泥产量

年份		2003	2004	2005	2006	2007	2008
水泥产量（万吨）		580	685	819	900	1 010	1 160
增长量（万吨）	逐期	—					

续表

年份		2003	2004	2005	2006	2007	2008
发展速度（%）	累计	—					
	环比	—					
	定基	100					
增长速度（%）	环比	—					
	定基	—					
增长 1%的绝对值		—					

（4）某商场 2002—2007 年销售额资料如表 6-24 所示。试计算表中空格的指标，并填入表中。

表 6-24　某商场 2002—2007 年销售额　　单位：万元

年份	发展水平	增减量		平均增减量	发展速度%		增减速度%		增减 1 的绝对值
		累计	逐期		定基	环比	定基	环比	
2002									
2003				42.5					2.85
2004		106.2							
2005							45.2		
2006						136.0			
2007								3.2	

（5）某地区甲产品 2005—2008 年各季收购量统计资料如表 6-25 所示。

要求：

1）用移动平均法对该时间数列进行修匀（列表表现其趋势值）。

2）用直接平均法计算其季节比率，并绘制季节变动图。

3）预计 2009 年该产品全年收购量为 96 万吨，按其季节比率，各季度的收购量应安排多少？

表 6-25　某地区甲产品 2005—2008 年各季收购量　　单位：万元

年份	一季度	二季度	三季度	四季度
2005	13	5	8	18
2006	14	6	10	18
2007	16	8	12	22
2008	19	15	17	25

CHAPTER 7

第 7 章　统计指数

章首引例

新华网北京 2009 年 1 月 22 日电（记者王希、刘铮）：国家统计局局长马建堂 22 日说，2008 年我国全年居民消费价格指数（CPI）比上年上涨 5.9%，涨幅比上年提高 1.1 个百分点。

在 2008 年国民经济运行情况新闻发布会上，马建堂介绍说，2008 年我国全年居民消费价格涨幅前高后低，下半年逐步回稳。全年居民消费价格指数比上年上涨 5.9%，涨幅比上年提高 1.1 个百分点。12 月单月 CPI 同比增长 1.2%。

分类别看，2008 年食品价格上涨 14.3%，拉动价格总水平上涨 4.65 个百分点；居住价格上涨 5.5%，拉动价格总水平上涨 0.82 个百分点；其余各类商品价格有涨有落。

2008 年工业品出厂价格比上年上涨 6.9%，12 月同比下降 1.1%，全年涨幅比上年提高 3.8 个百分点。原材料、燃料、动力购进价格上涨 10.5%，涨幅比上年提高 6.1 个百分点。70 个大中城市房屋销售价格比上年上涨 6.5%，12 月同比下降 0.4%，全年涨幅比上年回落 1.1 个百分点。

指数作为一种分析工具，不仅被广泛应用在分析社会经济发展预测上，而且还是对综合国力、生活水平、经济效益、金融运行状态等综合研究评判的重要工具。本章主要介绍指数概念和编制方法以及相关理论。

本章学习目标

- ☑ 理解统计指数的概念、含义、作用与种类；
- ☑ 熟练掌握综合指数、平均指数的编制原则和方法；
- ☑ 了解指数体系的含义，熟练掌握总量指标和平均指标的因素分析法；
- ☑ 了解常用的几种经济指数、指数数列的含义和编制方法。

7.1 统计指数及其种类

7.1.1 指数的概念

指数的概念有广义与狭义之分。从广义上讲，凡是说明社会经济现象数量变动的相对

数就是指数。如前面讲过的发展速度就可以称为指数。而从狭义上讲，统计指数是指那些不能直接加总和不能直接对比的复杂经济总体的综合变动方向与变动程度的相对数，是一种特殊的动态相对数。例如，要研究不同时期的钢材、煤炭、汽车、机床等多种产品产量的综合变动情况，由于它们各具有不同的使用价值和计量单位，不能在数量上直接加总，因而就无法将不同时期的总产量直接对比，就不能说明多种产品产量的综合变动方向和变动程度。这类不能直接加总的复杂的经济现象需要通过狭义指数来说明其综合变动情况。所以本章所阐述的统计指数就是这种狭义上的统计指数。

7.1.2　统计指数的作用

统计指数的作用主要有以下几个方面。

1．综合反映复杂现象总体数量上的变动方向和程度

运用统计指数，可以把不能直接加总的复杂现象的数量，过渡到可以进行相加和相互比较，以相对数的形式表明这种现象在数量上的综合变动方向和变动程度。在综合指数形式的编制指数中，根据其分子与分母的差额，还可以观察现象绝对量的增减变动情况。

2．分析现象总体变动中受各个因素变动的影响程度

任何复杂的社会经济现象总体的数量变动都是其构成因素的变动综合影响的结果。如工业产值的变动，不仅受工业产品产量多少的影响，而且还要受其价格高低的影响，二者共同影响的结果表现为工业产值的总变动。通过统计指数方法，计算出现象总体和各种因素指数，就可以深入分析各因素对总体变动的影响程度和绝对额。

3．社会经济现象在长时期内的发展变化趋势

将同类指数按时间的先后顺序排列编制成动态数列，可以分析和研究社会经济现象在长时期内的发展变化趋势。

分析指数的这三个方面的作用将在下面各节中做详细说明。

7.1.3　统计指数的种类

统计指数可以从不同的角度进行分类。

1．统计指数按研究的范围不同，分为个体指数和总指数

（1）个体指数

个体指数是反映总体中个别现象或个别项目变动的相对数。例如，反映个别产品产量变动的个体产量指数，反映个别商品价格变动的个体价格指数等。

其计算公式为：

$$\text{个体指数}=\frac{\text{报告期水平}}{\text{基期水平}} \tag{7-1}$$

设：q 表示产品产量，p 表示产品价格，k 表示个体指数，则个体产量指数为：

$$k_q = \frac{q_1}{q_0}$$

个体价格指数为：

$$k_p = \frac{p_1}{p_0}$$

（2）总指数

总指数是反映整个总体现象数量综合变动的特殊相对数。例如，反映多种产品的总产量变动的产量总指数，反映多种商品价格总变动的价格总指数。总指数用代号 $\bar{k}$ 表示，如产量指数 $\bar{k}_q$，价格总指数 $\bar{k}_p$。总指数的编制方法将在 7.2 节中详细介绍。

2．统计指数按说明的现象的性质不同，分为数量指标指数和质量指标指数

- 数量指标指数是反映复杂现象总体在规模、水平上数量变动程度的指数，如工业产品产量指数，商品销售量指数等。
- 质量指标指数是反映复杂现象总体在内涵上数量变动程度的指数，如价格指数、成本指数、劳动生产率指数等。

3．统计指数按所采用的基期不同，分为环比指数和定基指数

- 环比指数是指在一个指数数列中，各时期的指数都是以前一期为基期所编制的。环比指数反映某种现象逐期变动的趋势和变动程度。
- 定基指数是指在一个指数数列中，各时期的指数都是按照某一固定的基期所编制的。定基指数反映某种现象在一个较长时期内的变动趋势和变动程度。

4．总指数按编制的方法不同，分为综合指数和平均指数

- 综合指数是指通过同度量因素，将两个时期不能同度量的现象指标过渡到能够同度量的指标，然后再计算出的指数。它是总指数编制的基本形式。
- 平均指数是从个体指数出发，通过对个体指数加权平均计算而编制的指数。这是总指数的另一种计算形式。

7.2 总指数的编制与应用

总指数按编制的方法不同，分为综合指数法和平均指数法两种。下面就分别介绍这两种指数的编制及其应用。

7.2.1　综合指数法

1．综合指数的意义

综合指数是由两个不同时期的总量指标进行对比的，指标中包含两个或两个以上的因素，我们将其中被研究因素以外的一个或一个以上的因素固定下来，仅观察被研究因素的变动情况，这样编制出来的总指数叫综合指数。它的特点是先综合后对比。

2．综合指数的编制原理

首先，引入同度量因素，以解决复杂总体所包括的许多个别事物在研究指标上不能直接相加的问题。为此，需要从现象的内在联系出发，确定与研究现象相联系的因素，使其成为同度量因素，从而将不能直接加总的指标过渡到能够相加和对比的指标，然后进行对比。所谓同度量因素，是使不能直接相加的不同度量现象过渡到可以加总的那个因素。例如，不同商品的销售量由于经济意义不同，不能直接相加。如果销售量乘以价格即成为销售额，而销售额是价值指标，具有广泛的综合性，可以相加和对比。这里我们就利用了价格这一同度量因素。

其次，将同度量因素加以固定，以消除同度量因素变动的影响，测定所要研究的因素的变动情况。例如，我们要研究的是两个时期产品销售总额中各类产品销售量的变动，就需要将作为同度量因素的产品价格固定在同一个时期，以测定两个时期各类产品销售量的变动情况。

最后，将两个时期的总量指标对比，其结果既为综合指数，也就综合地反映了复杂总体研究指标的变动情况。

在实际应用中，将同度量因素固定在报告期还是基期是个十分重要的问题，因为同度量因素不仅起同度量的作用，而且具有加权的作用，用不同时期的同度量因素计算，会得到不同的综合指数结果。所以，编制数量指标指数和质量指标指数时，各有不同的解决方法和规定。下面分别阐述。

3．数量指标综合指数的编制及应用

在编制数量指标指数以对数量指标的综合变动进行分析时，需要确定相应的同度量因素。

下面以产量总指数为例说明数量指标指数的编制方法。不同的产品有不同的计量单位，在数量上不能直接相加，必须通过同度量因素将不可加总的产量还原成可以相加的产值。在还原的过程中，价格充当了同度量因素的角色。从产量过渡到产值后要使在产值的变动中体现产量的变动，就要对作为同度量因素的价格加以固定，以排除价格变动的影响。

设 q 表示产量，p 表示价格，则有

$$\text{产量} \times \text{价格} = \text{产值}$$

即

$$q \times p = qp$$

$$\text{产量总指数} = \frac{\text{某一价格计算的报告期产值总和}}{\text{同一价格计算的报告期产值总和}}$$

用 $\bar{k}$ 表示总指数，则产量总指数 $\bar{k}_q$ 的计算公式为：

$$\bar{k}_q = \frac{\sum q_1 p}{\sum q_0 p} \tag{7-2}$$

现在的问题就是如何固定价格 p 这个同度量因素。用什么时期的价格，有以下两种选择。

- 选择基期价格作为同度量因素，则

$$\bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

式中，$\sum q_1 p_0$ 是按基期价格计算的报告期产值之和；$\sum q_0 p_0$ 是基期产值之和。

- 选择报告期价格作为同度量因素，则

$$\bar{k}_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

式中，$\sum q_1 p_1$ 是报告期产值之和；$\sum q_0 p_1$ 是按报告期价格计算的基期产值之和。



把同度量因素固定在基期来编制综合指数的方法，是在 1864 年由德国经济学家埃蒂恩·拉斯贝尔（Etienne Laspeyres）首先提出来的。因而将其称为拉氏公式。他主张不论是数量指标指数还是质量指标指数都采用基期同度量因素（权数）的指数。如以上第一种选择。

把同度量因素固定在报告期来编制综合指数的方法，是由德国经济学家哈曼·派许（Hermann Paasche）在 1874 年首先提出的，故称为派许公式，简称为派氏公式。他主张不论是数量指标指数还是质量指标指数都采用报告期同度量因素（权数）的指数。如以上第二种选择。

对于以上两种情况，在实际工作当中，一般选基期价格作为同度量因素，即选用拉氏计算公式编制数量指标综合指数。这是因为：首先，编制产量指数的目的是要排除价格因素变动的影响，单纯反映产品产量的总变动，若把价格固定在报告期，这就包含了价格由过去到现在的变化因素，而不能真实地反映产量的变动。要使价格不变，就应该用基期价格，才符合经济现象变动的客观实际。其次，应满足构成指数体系的要求。

所以，编制数量指标综合指数的原则是：在编制数量指标综合指数时，应将作为同度量因素的质量指标固定在基期。

数量指标综合指数的计算公式应为:

$$\bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \tag{7-3}$$

在统计工作的实践中，为了研究比较长时期的产量变动情况，便于对各期产量指数进行比较，在编制工业产品产量指数时，而把价格固定在某个更长远的基期上，即采用不变价格作为同度量因素。其计算公式为：

$$\bar{k}_q = \frac{\sum q_1 p_n}{\sum q_0 p_n} \tag{7-4}$$

式中，p_n 是不变价格。

【例 7-1】

说明数量指标综合指数的编制过程及其数值的实际意义，如表 7-1 所示。

表 7-1　某工业企业三种产品的产量及价格资料

产品名称	计量单位	产量 q		价格 p（元）	
		基期 q_0	报告期 q_1	基期 p_0	报告期 p_1
甲	吨	1 200	1 440	50	59
乙	件	600	660	80	92
丙	米	3 000	3 450	25	27

从上表中可看出，三种产品产量都有所变动。我们要研究三种产品产量总的变动情况，就需要编制三种产品产量的综合指数 $\bar{k}_q$：

$$\bar{k}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{1\,440\times 50 + 660\times 80 + 3\,450\times 25}{1\,200\times 50 + 600\times 80 + 3\,000\times 25} = \frac{211\,050}{183\,000} = 1.15\,33 = 115.33\%$$

绝对数：　$\sum q_1 p_0 - \sum q_0 p_0 = 211\,050 - 183\,000 = 28\,050$（元）

计算结果表明，该企业三种工业产品产量报告期比基期总的增长了 15.33%，产量的提高使得产值报告期比基期增加了 28 050 元。

4．质量指标综合指数的编制与应用

在编制质量指标指数以对质量指标的综合变动进行分析时，也要确定相应的同度量因素。下面以价格总指数为例来说明质量指标指数的编制方法。

不同产品的价格虽然在形式上都用货币作为计量单位，但它仍然属于不能直接加总的非同度量现象。在数值上仍然不是有可加性，必须通过同度量因素将不可加总的价格还原为可以加总的产值。在这还原的过程中，产量就成了同度量因素。

仍设 q 为产量，p 为价格，则有

$$\text{价格} \times \text{产量} = \text{产值}$$

即

$$p \times q = pq$$

q 作为同度量因素，在计算时，仍需要固定在同一个时期。

用 $\bar{k}_p$ 表示价格综合指数，则其计算公式为：

$$\text{价格总指数} = \frac{\text{某一时期产量计算的报告期产值之和}}{\text{同一时期产量计算的基期产值之和}}$$

$$\bar{k}_p = \frac{\sum qp_1}{\sum qp_0} \qquad (7\text{-}5)$$

现在的问题仍然是选择 q 的时期的问题。用什么时期的产量，同样有以下两种选择。

- 选择基期产量作为同度量因素，则产量综合指数采用拉氏计算公式：

$$\bar{k}_p = \frac{\sum q_0 p_1}{\sum q_0 p_0}$$

式中，$\bar{k}_p$ 是质量指标综合指数；$\sum q_0 p_1$ 是按基期价格计算的报告期产值之和；$\sum q_0 p_0$ 是基期产值之和。

- 选择报告期产量作为同度量因素，则产量综合指数采用派氏计算公式：

$$\bar{k}_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}$$

式中，$\sum q_1 p_1$ 是报告期产值之和。

对于以上两种情况，在实际工作中，一般选择报告期产量为同度量因素，即采用派氏公式编制产量指标指数。这是因为：一是从现实的经济意义考虑，人们所关心的是因为价格水平变化而使现在生产的产品增加或减少的产值，而不是关心过去生产的产品增加或减少的产值。所以只有用报告期的产量才能反映出由于价格变化而使报告期产品价格增加或减少的实际数额。二是选择报告期产量作为同度量因素，可满足构成指数体系的要求。

所以，编制质量指标综合指数的原则是：在编制质量指标综合指数时，应将作为同度量因素的数量指标固定在报告期。

质量指标综合指数的计算公式应为：

$$\bar{k}_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} \qquad (7\text{-}6)$$

➪【例 7-2】

现仍以表 7-1 资料为例，说明质量指标综合指数的编制过程及其应用。

根据表 7-1 的资料，三种产品的价格均有所变动，要反映三种产品的价格的总变动情况，必须应用综合指数的计算公式。

$$\bar{k}_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{1\,440\times 59 + 660\times 92 + 3\,450\times 27}{1\,440\times 50 + 660\times 80 + 3\,450\times 25} = \frac{238\,830}{211\,050} = 1.131\,6 = 113.16\%$$

绝对数：$\sum q_1 p_1 - \sum q_1 p_0 = 238\,830 - 211\,050 = 27\,780$ (元)

计算结果表明，该企业三种产品价格报告期比基期总的增长了 13.16%，使产值报告期比基期增加了 27 780 元。

7.2.2　平均指数法

从前面介绍的综合指数的编制可看出，在计算综合指数时，需要较多的数据资料。但在实际的工作中，经常会出现资料不全或只有间接资料的情况。比如，只有个体价格指数，而缺乏基期或报告期价格；只有基期或报告期的产值，而缺乏基期与报告期产量等。这时，就需要采用平均指数法计算总指数。平均指数法是以个体指数为基础，采用加权形式来编制的总指数。根据选用的权数不同，平均指数法可进一步分为加权算术平均法、加权调和平均法、固定权数加权平均法。

1．加权算术平均法

加权算术平均法是以个体指数为变量值，以一定时期的总值资料为权数，对个体指数加权平均以计算总指数的方法。

➪【例 7-3】

某企业有关生产情况如表 7-2 所示。要求根据以下资料反映该企业产量总的变动情况。

表 7-2　某企业生产情况

产品名称	基期产值（万元）	个体产量指数(元)
甲	280	112
乙	370	108
丙	350	114
合计	1 000	—

若按数量指标综合指数 $\bar{k}_q=\frac{\sum q_1p_0}{\sum q_0p_0}$ 公式计算，则缺乏分子资料，不能直接应用，而应做适当的变形，将 $k_q=\frac{q_1}{q_0}\Rightarrow q_1=k_qq_0$ 代入 $\bar{k}_q$ 的综合指数计算公式，得

$$\bar{k}_q=\frac{\sum q_1p_0}{\sum q_0p_0}=\frac{\sum k_qq_0p_0}{\sum q_0p_0} \qquad (7\text{-}7)$$

式中，k_q 是个体产量指数，是变量值；q_0p_0 是基期产值，是权数。

$\bar{k}_q=\frac{\sum q_1p_0}{\sum q_0p_0}$ 在表现形式上同加权算术平均数计算公式 $\bar{x}=\frac{\sum xf}{\sum f}$ 一样，故称为加权算术平均指数。将表中数据代入式（7-7），得

$$\bar{k}_q=\frac{\sum k_qq_0p_0}{\sum q_0p_0}=\frac{1.12\times280+1.08\times370+1.14\times350}{1\,000}=\frac{1\,112.2}{1\,000}=1.112\,2=111.22\%$$

绝对数：$\sum k_qq_0p_0-\sum q_0p_0=1\,112.2-1\,000=112.2$（万元）

计算结果表明，该企业三种产品产量报告期比基期平均增长了 112.2%。由于产量变动，使企业总产值报告期比基期增加了 112.2 万元。

2．加权调和平均法

➪【例 7-4】

根据表 7-3 所示的企业生产情况来求该企业产品价格的变动情况。

表 7-3　某企业生产情况

产品名称	报告期产值（万元）	个体价格指数（%）
甲	330	109
乙	430	111

续表

产品名称	报告期产值（万元）	个体价格指数（%）
丙	440	113
合计	1 200	—

若按质量指标综合指数 $\bar{k}_p=\dfrac{\sum q_1p_1}{\sum q_1p_0}$ 公式计算，则缺乏分母资料，不能直接计算，应做如下变形，将 $k_p=\dfrac{p_1}{p_0}\Rightarrow p_0=\dfrac{p_1}{k_p}$ 代入 $\bar{k}_p$ 的综合指数计算公式，得

$$\bar{k}_p=\frac{\sum q_1p_1}{\sum q_1p_0}=\frac{\sum q_1p_1}{\sum q_1\dfrac{p_1}{k_p}}=\frac{\sum q_1p_1}{\sum\dfrac{1}{k_p}q_1p_1} \tag{7-8}$$

式中，k_p 是个体价格指数，是变量值；q_1p_1 是报告期产值，是权数。

$\bar{k}_p=\dfrac{\sum q_1p_1}{\sum\dfrac{1}{k_p}q_1p_1}$ 在表现形式上同调和平均数计算公式 $h=\dfrac{\sum m}{\sum\dfrac{1}{x}m}$ 一样，故将其称为加权调和平均指数。

将表中数据代入式（7-8），得

$$\bar{k}_p=\frac{\sum q_1p_1}{\sum\dfrac{1}{k_p}q_1p_1}=\frac{1\,200}{\dfrac{330}{1.09}+\dfrac{430}{1.11}+\dfrac{440}{1.13}}=\frac{1\,200}{1\,079.52}=1.111\,6=111.16\%$$

绝对数：$\sum q_1p_1-\sum\dfrac{1}{k_p}q_1p_1=1\,200-1\,079.52=120.48$（万元）

计算结果表明，该企业三种产品价格报告期比基期平均增长了 11.16%；价格变动使企业总产值报告期比基期增加了 120.48 万元。

3．固定权数平均法

（1）固定权数平均指数的概念

前面所介绍的加数算术平均数指数和加数调和平均数指数计算公式中的权数都是以绝对数形式出现的，要求有全面的统计资料，而在实际工作中往往很难满足。因此，在统计实践中，常常把这个权数用比重的形式固定下来，一段时间内不做变动，这种权数称为固定权数。以指数化因素的个体指数为基础，使用固定权数对个体指数（或类指数）进行加权平均计算的总指数称为固定权数平均指数。

固定权数用 w 表示，$\sum w=100$，固定权数平均数指数有以下计算公式：

$$\bar{k}=\frac{\sum kw}{\sum w}$$ 称为固定权数加权算术平均指数 （7-9）

$$\bar{k}=\frac{\sum w}{\sum \frac{1}{k}w}$$ 称为固定权数加权调和平均数指数 （7-10）

采用固定权数的加权平均指数，不仅可以避免每次编制指数权数资料来源的困难，而且也便于前后不同时期比较。

（2）固定权数平均指数的应用

世界上大多数国家都编制居民消费价格指数（CPI），反映城乡居民购买并用于消费的消费品及服务价格水平的变动情况，并用它来反映通货膨胀程度。从 2001 年起，我国采用国际通用做法，逐月编制并公布以 2000 年价格水平为基期的居民消费价格定基指数，作为反映我国通货膨胀（或紧缩）程度的主要指标。下面以我国居民消费价格指数为例，说明固定权数加权平均法的应用。

第一步：将全部消费品及服务项目分类。居民消费价格指数包括居民用于日常生活的全部商品和服务项目。按国家统计局《居民消费价格指数商品及服务项目目录》规定共分 8 个大类，即食品、烟酒及用品、衣着、家庭设备用品及维修服务、医疗保健及个人用品、交通和通信、娱乐教育文化用品及服务、居住。每个大类包括若干个中类，中类之下又有基本分类。根据全国城乡 9 万余户居民家庭消费支出调查资料中消费额较大的项目和习惯确定，共设 351 个基本分类。再在各基本类中选出若干代表规格品。

代表规格品是按照消费量较大，其价格变动趋势和变动程度有较强代表性的合格品的要求选择的，并且都规定了最低数量标准，各地可根据当地实际情况适当增加。

第二步：计算各类消费品及服务项目的权数。根据统计调查和统计报表资料结合社会商品零售额统计资料，计算各类商品零售额在社会商品零售额中所占的比重，将其作为权数并固定下来，若干年保持不变。为连续编制物价指数提供了有利条件。

第三步：分层计算各分类指数。先以代表规格品个体物价指数计算基本分类物价指数，将计算出的基本分类指数作为个体物价指数计算中类物价指数，以此类推，计算大类物价指数，最终由各大类指数加权为居民消费价格总指数。基本分类指数是用简单几何平均法对若干代表规格品的个体物价指数进行平均，中类和大类及总指数则均采用固定权数加权算术平均法。

➪【例 7-5】

现以各大类物价指数资料计算物价指数的计算过程，来说明固定权数加权计算平均指数在居民消费价格总指数编制过程中的应用，如表 7-4 所示。

表 7-4　某地区居民消费价格总指数计算表

商品分类	物价指数 k_p（%）	固定指数 w（%）	k_pw
食品类	106.3	390	41.60
烟酒及用品类	100.7	30	3.02
衣着类	103.8	70	7.27
家庭设备用品及维修服务类	100.9	80	8.07
医疗保健及个人用品类	99.8	90	8.98
交通和通信类	99.4	75	7.46
娱乐教育文化用品及服务类	101.2	152	15.38
居住类	98.5	113	11.13
合　计	102.8	1 000	10 285.00

该地区居民消费价格总指数为：

$$\bar{k}=\frac{\sum k_p w}{\sum w}=\frac{10\,285}{1\,000}=102.8\%$$

相关链接

指数在社会经济统计中应用很广泛。在我国统计实践中，重要的统计指数有如下几种：

（1）商品零售物价指数

商品零售物价指数是反映城乡商品零售价格变动趋势的一种经济指数，其计算方法和过程与居民消费价格指数相似，主要的不同在于它不能反映居住和服务价格的变动。从2001年开始，我国价格统计指数的公布和使用由以商品零售价格指数为主改为以居民消费价格指数为主，与国际惯例接轨。

（2）农产品收购价格指数

农产品收购价格指数是反映国家农产品收购价格变动趋势和程度的指数，是研究农产品收购价格变化对农民收入、国家财政支出等的影响，又是计算工农业产品综合比价指数的依据。它是以报告期农产品实际收购金额为权数，对各类代表规格品个体价格指数采用加权调和平均法计算农产品收购价格总指数，其计算公式为：

$$\bar{k}_p=\frac{\sum p_1q_1}{\sum\frac{1}{k_p}p_1q_1}$$

（3）工业生产指数

工业生产指数概括反映了一个国家或地区各种工业产品产量的综合变动程度，它是衡量经济增长水平的重要指标之一。世界各国都非常重视工业生产指数的编制，但采用的方法完全不同。我国工业生产指数是通过计算各种工业产品的不变价格产值来编制的。其计算公式为：

$$\bar{k}_p = \frac{\sum p_n q_t}{\sum p_n q_0}$$

或 $$\bar{k}_p = \frac{\sum p_n q_t}{\sum p_n q_{t-1}}$$

国外较为普遍采用平均指数的形式来编制工业生产指数，其计算公式为：

$$\bar{k}_p = \frac{\sum k_q p_0 q_0}{\sum p_0 q_0}$$

为了简化指数编制工作，常以各种工业品的增加值比重作为权数，并且将其相对固定起来，即运用固定加权算术平均指数法，连续地编制各个时期的工业生产指数，与居民消费价格指数的编制相似。

（4）股票价格指数

在发育比较充分的市场经济条件下，股票价格的波动和走向是反映经济景气状况的重要方面，也是影响投资人决策行为的主要因素之一。股票价格指数（简称股价指数）可以衡量整个股票市场价格变动的基本趋势，人们形象地称之为市场经济的“晴雨表”。

股票价格指数的计算方法很多，但一般以发行量为权数进行加权综合，其计算公式为：

$$\bar{k}_{1/0} = \frac{\sum p_{1i} q_i}{p_{0i} q_i}$$

式中，p_{1i} 为第 i 种股票报告期价格，p_{0i} 为第 i 种股票基期价格，q_i 为第 i 种股票的发行量，大多固定在报告期。

7.3 指数体系与因素分析

7.3.1 指数体系

1. 指数体系的概念

社会经济现象客观上存在着错综复杂的经济联系，这种经济联系，从静态上分析，表

现为某些社会经济现象之间具有数量上的必然联系，如：

产品产值=产品产量×产品价格

商品销售额=商品销售量×商品价格

对这种在静态上的数量关系，按指数形式表现时，同样也存在这种对等关系，即：

产品产值总指数=产品产量总指数×产品价格总指数

商品销售额总指数=商品销售量总指数×商品价格总指数

在统计研究中，把由这样三个或三个以上具有内在联系的，即经济上有联系、在数量上保持一定对等关系的统计指数组成的整体，称为指数体系。

指数体系内在的数量对等关系也表现在绝对数之间，即：

产品产值的增减额=产品产量变动影响的产值增减额+产品价格变动影响的产值增减额

商品销售额的增减额 = 商品销售量变动影响的销售额增减额 ×
商品价格变动影响的销售额增减额

上述指数体系，按编制综合指数的一般原理，可写成：

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \tag{7-11}$$

$$(\sum q_1 p_0 - \sum q_0 p_0) + (\sum q_1 p_1 - \sum q_1 p_0) = \sum q_1 p_1 - \sum q_0 p_0 \tag{7-12}$$

$$\text{产品产量总指数} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

$$\text{产品价格总指数} = \frac{\sum q_1 p_1}{\sum q_1 p_0}$$

$$\text{产品产值总指数} = \frac{\sum q_1 p_1}{\sum q_0 p_0}$$

2．指数体系的作用

（1）利用指数体系可以进行指数间的推算

指数体系表现为一个数量对等关系式，可以依据其组成的体系等式，根据已掌握的若干个指数，推算出体系中的某一个未知指数。

例如：某厂产量计划提高 10%，单位成本计划降低 5%，总成本将如何变动?

根据已知资料得知：产量指数为 110%，单位成本指数为 95%。

按指数体系为：

总成本指数=产量指数×单位成本指数=110%×95%=104.5%

表明：总成本将提高 4.5%。

（2）可以对复杂社会经济现象的总变动进行因素分析

指数体系既反映了社会经济现象的总变动，又反映了其中各因素的变动。因而，可以利用指数体系对现象的总变动进行影响因素的定量分析，确定各因素变动对现象总变动在方向和程度上的影响程度，其中包括两方面的分析：相对数分析和绝对数分析。如产品产值总指数=产品产量总指数×产品价格总指数，是从相对数方面分析；产品产值的增减额=产品产量变动影响的产值增减额+产品价格变动影响的产值增减额，则是从绝对数的角度分析。指数体系是因素分析的依据。

7.3.2 因素分析

1. 因素分析的含义

因素分析是利用指数体系从数量方面研究现象的综合变动中，各个因素变动对其影响的方向、程度和绝对效果的方法，这一方法也被称为指数分析法。

因素分析法的研究对象是复杂的经济现象，这些复杂现象受两个或两个以上因素变动的共同影响。因素分析的目的，就是测定各个因素对总体的影响方向和程度。

因素分析法的基本依据是指数体系。现象总变动指数等于若干因素指数的乘积，现象总变动的差额等于若干因素影响差额的总和，这是因素分析法计算的依据。

因素分析法的基本特点是以假定一个因素变动、其余因素不变为前提的。如果是三个因素的影响，必须假定其中两个因素不变，只测定另一个因素的影响。依因素分步进行，每一步只测定诸因素中的一个因素的影响方向和程度。

因素分析法的作用是揭示复杂现象总变动和影响复杂现象变动的各种因素变动的相对数和绝对数的变动方向与程度，为经济活动分析和决策提供可靠依据。

2. 因素分析的种类

因素分析按被研究指标的性质不同可分为两类：总量指标的因素分析和平均指标的因素分析。前者分析绝对指标的总变动中各因素变动的影响，如总产值变动中的价格、产量的影响情况；后者分析加权平均指标的变动中变量值水平和总体结构的影响，如平均工资的变动中工资水平和工人结构的影响情况等。

因素分析按分析时包含的因素多少可分为两类：两因素分析和多因素分析。两因素分析是指研究对象仅包含两个因素的变动的分析，它是因素分析的基本方法，如销售额受销售价格和销售量的影响分析。多因素分析是指研究对象包含有两个以上的因素变动的分析，如原材料支出额受产量、原材料单耗、原材料价格的影响。

3．因素分析的步骤

因素分析法的运用应建立指数体系，并依据指数体系从相对数及绝对数两个方面进行分析计算。其步骤是：首先，计算被分析指标的总变动程度和绝对数。其次，计算各因素变动影响程度和影响绝对数。最后，计算指数体系间的等量关系并进行综合分析。

7.3.3　总量指标因素分析

总量指标因素分析是指对复杂现象总量指标和总量指标分解后各类因素的分析。例如，产值受产量、出厂价格因素的影响分析，销售额受销售量和销售价格因素影响的分析。

1．两因素分析

两因素分析是指一个现象总变动受两个因素影响时，分析其中每个因素的变动对总变动影响的方向和程度。

⇨【例 7-6】

现以表 7-5 的资料，分析某商场三种商品销售额的变动，受销售量因素和价格因素变动的影响程度。

表 7-5　某商场三种商品销售资料

商品名称	计量单位	销售量 q		价格 p (元)	
		基期 q_0	报告期 q_1	基期 p_0	报告期 p_1
甲	套	300	380	280	260
乙	件	450	510	140	150
丙	米	1 200	1 500	70	78

（1）计算销售额总指数，测定现象总体变动的程度和绝对额

$$\bar{k}_{qp}=\frac{\sum q_1p_1}{\sum q_0p_0}=\frac{380\times260+510\times150+1\,500\times78}{300\times280+450\times140+1\,200\times70}=\frac{292\,300}{231\,000}=1.265\,3=126.54\%$$

即：销售额报告期比基期增长了 26.54%。

销售额绝对值变动：　$\sum q_1p_1-\sum q_0p_0$ =292 300 – 231 000 = 61 300（元）

（2）分别计算销售量和单价两个因素变动的影响程度和绝对额

销售量总指数：

$$\bar{k}_{q}=\frac{\sum q_1p_0}{\sum q_0p_0}=\frac{380\times280+510\times140+1\,500\times70}{300\times280+450\times140+1\,200\times70}=\frac{282\,300}{231\,000}=1.224\,2=122.42\%$$

绝对额：　$\sum q_1p_0-\sum q_0p_0$ = 282 800—231 000=51 800（元）

这表明该商场由于销售量的变动，使得销售额报告期比基期增长了22.42%，销售额增加了518 000元。

销售价格总指数:

$$\bar{k}_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{292\,300}{282\,800} = 1.033\,6 = 103.36\%$$

绝对额: $\sum q_1 p_1 - \sum q_1 p_0 = 292\,300 - 282\,800 = 9\,500$（元）

这表明该商场由于销售价格的变动，使得销售额报告期比基期增长了3.36%，销售额增加了9 500元。

（3）利用指数体系从相对数和绝对数两方面进行影响因素的综合分析

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0}$$

即: $122.42\% \times 103.36 = 126.53\%$

绝对额: $(\sum q_1 p_0 - \sum q_0 p_0) + (\sum q_1 p_1 - \sum q_1 p_0) = \sum q_1 p_1 - \sum q_0 p_0$

$= 51\,800 + 9\,500 = 61\,300$（元）

计算和分析结果表明:

该商场销售额报告期比基期增长了26.54%，销售额增加了61 300元，是由于销售量的变动，使销售额增长了22.42%，绝对额增加了51 800元，和由于销售价格的变动，使销售额增长了3.36%，绝对额增加了9 500元的共同结果。

2．多因素分析

多因素分析是指一个现象的变动受多个（三个以上）因素影响时分析其中每个因素的变动对总变动影响的方向和程度。如原材料费用总额受产品产量、单位产品原材料消耗量和单位原材料价格三个因素的影响。

在静态情况下，四者的关系为:

原材料费用总额=产品产量×单位产品原材料消耗量×单位原材料价格

那么在动态的情况下四者之间的关系如何?

首先编制四个综合指数。设产品产量仍为 q，单位产品原材料消耗量为 m，单位原材料价格为 p。

根据因素分析的原则，在测定某一因素指标变动影响时将其他因素作为同度量因素固定起来，同度量因素为数量指标的固定在报告期，同度量因素为质量指标的固定在基期。

但在编制多因素综合指数时应注意一个问题，就是在区分数量因素和质量因素时，由



于分析的因素较多，要注意应相对地而不能绝对地判断。如上例中，产量（q）是数量指标，而单位产品消耗原材料的金额（mp）相对产量来说是质量指标。因此分析产量变动影响时，要把 mp 固定在基期。单位产品原材料消耗量（m）相对产量（q）是质量指标，相对原材料单价（p）是数量指标。所以分析单位产品原材料消耗量的变动影响时，则要把产量固定在报告期，原材料单价固定在基期。对于原材料单价（p），相对原材料消耗总量（qm）这一数量指标来说，它是质量指标。因此分析其对原材料费用总额的变动影响时，应将 qm 固定在报告期。

根据以上原则编制各个综合指数：

$$\text{产品产量综合指数}\quad \bar{k}_q=\frac{\sum q_1m_0p_0}{\sum q_0m_0p_0}$$

$$\text{单位产品原材料消耗量综合指数}\quad \bar{k}_m=\frac{\sum q_1m_1p_0}{\sum q_1m_0p_0}$$

$$\text{单位原材料价格综合指数}\quad \bar{k}_p=\frac{\sum q_1m_1p_1}{\sum q_1m_1p_0}$$

$$\text{原材料费用总额指数}\quad \bar{k}_{qmp}=\frac{\sum q_1m_1p_1}{\sum q_0m_0p_0}$$

四个指数的关系：

$$\frac{\sum q_1m_0p_0}{\sum q_0m_0p_0}\times\frac{\sum q_1m_1p_0}{\sum q_1m_0p_0}\times\frac{\sum q_1m_1p_1}{\sum q_1m_1p_0}=\frac{\sum q_1m_1p_1}{\sum q_0m_0p_0}\qquad(7\text{-}13)$$

即：原材料费用总额指数=产品产量综合指数×单位产品原材料消耗量综合指数×单位原材料价格综合指数

➪【例 7-7】

假设某企业生产产品有关资料如表 7-6 所示，要求运用指数体系分析产品产量、单位产品原材料消耗量及原材料价格对原材料费用总额的影响。

表 7-6　某企业生产及原材料消耗资料

产品名称	计量单位	产品产量		单位原材料消耗量		单位原材料价格（元）	
		基期 q_0	报告期 q_1	基期 m_0	报告期 m_1	基期 p_0	报告期 p_1
甲	台	100	120	10	11	20	22
乙	件	300	360	8	7	15	14
丙	米	1 000	1 500	5	4	8	7

进行如下计算：

（1）原材料费用总额指数

$$\bar{k}_{qmp}=\frac{\sum q_1m_1p_1}{\sum q_0m_0p_0}=\frac{120\times11\times22+360\times7\times14+1\,500\times4\times7}{100\times10\times20+300\times8\times15+1\,000\times5\times8}=\frac{106\,320}{96\,000}=1.107\,5=110.75\%$$

绝对数：$\sum q_1m_1p_1-\sum q_0m_0p_0$ =106 320－96 000=10 320（元）

计算表明该企业原材料费用总额报告期比基期增长了 10.75%，费用总额增加了 10 320 元。

（2）产品产量指数

$$\bar{k}_q=\frac{\sum q_1m_0p_0}{\sum q_0m_0p_0}=\frac{120\times11\times20+360\times8\times15+1\,500\times5\times8}{100\times10\times20+300\times8\times15+1\,000\times5\times8}=\frac{127\,200}{96\,000}=1.325=132.50\%$$

绝对数：$\sum q_1m_0p_0-\sum q_0m_0p_0$ =127 200—96 000=31 200（元）

表明该企业由于产品产量变动，使原材料费用总额报告期比基期增长了 32.5%，绝对额增加了 31 200 元。

（3）单位产品原材料消耗量指数

$$\bar{k}_m=\frac{\sum q_1m_1p_0}{\sum q_1m_0p_0}=\frac{120\times11\times20+360\times7\times15+1\,500\times4\times8}{100\times10\times20+300\times8\times15+1\,000\times5\times8}=\frac{112\,200}{127\,200}=0.882\,1=88.21\%$$

绝对数：$\sum q_1m_1p_0-\sum q_1m_0p_0$ =112 200—127 200=—15 000（元）

表明该企业由于单位产品原材料消耗量变动，使原材料费用总额报告期比基期下降了 11.79%，绝对额减少了 15 000 元。

（4）单位原材料价格指数

$$\bar{k}_p=\frac{\sum q_1m_1p_1}{\sum q_1m_1p_0}=\frac{106\,320}{112\,200}=0.947\,6\%$$

绝对数：$\sum q_1m_1p_1-\sum q_1m_1p_0$ = 106 320—112 200=—5 880（元）

表明：该企业由于单位原材料价格变动，使原材料费用总额报告期比基期下降了 5.24%，绝对额减少了 5 880 元。

（5）利用指数体系分析四者之间的数量关系

$$\frac{\sum q_1m_0p_0}{\sum q_0m_0p_0}\times\frac{\sum q_1m_1p_0}{\sum q_1m_0p_0}\times\frac{\sum q_1m_1p_1}{\sum q_1m_1p_0}=\frac{\sum q_1m_1p_1}{\sum q_0m_0p_0}$$

即：$132.50\%\times88.21\%\times94.76\%=110.75\%$

绝对数：$(\sum q_1 m_0 p_0 - \sum q_0 m_0 p_0) + (\sum q_1 m_1 p_0 - \sum q_0 m_0 p_0) + (\sum q_1 m_1 p_1 - \sum q_0 m_1 p_0)$

$= \sum q_1 m_1 p_1 - \sum q_0 m_0 p_0 = 31\,200 + (-15\,000) + (-5\,880) = 10\,320$（元）

整个计算分析表明：该企业原材料费用总额报告期比基期增长了 10.75%，增加了 10 320 元，是由于产品产量的变动使原材料费用总额增长了 32.50%，增加了 31 200 元；由于单位产品原材料消量变动，使原材料费用总额下降了 11.79%，绝对额减少了 15 000 元以及由于单位原材料价格变动使原材料费用总额下降了 5.24%，绝对额减少了 5 880 元共同作用的结果。

7.3.4　总平均指标因素分析

从静态的角度上来看，总体某一标志的平均水平的高低，既取决于各单位标志值水平的高低，也取决于标志值水平不同的总体单位的分布状况。

下面我们以平均工资为例，从动态的角度上来分析总平均工资的变动，受哪些因素的影响。

根据综合指数的编制原则，还可编制以下两个指数：

由　　$\bar{x} = \dfrac{\sum xf}{\sum f}$

得基期总平均工资　　$\bar{x}_0 = \dfrac{\sum x_0 f_0}{\sum f_0}$

报告期总平均工资　　$\bar{x}_1 = \dfrac{\sum x_1 f_1}{\sum f_1}$

那么，总平均工资的变动则可用指数 $\dfrac{\bar{x}_1}{\bar{x}_0}$ 表示，也即

$$\frac{\dfrac{\sum x_1 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_0}{\sum f_0}}$$

把这种指数称为可变构成指数。

根据综合指数的编制原则，还可编制以下两个指数：

一是反映各组标志值（x）变动的指数：

$$\frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_1}{\sum f_1}}$$

这种以各组水平变动为基础，将各组在总体内部的结构给予固定，以说明各组水平变动情况的指数称为固定结构指数。

二是反映总体内部各组结构变动的指数：

$$\frac{\frac{\sum x_0 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}}$$

这种以总体内部各组结构变动为基础，将各组水平给予固定，以说明总体内部各组结构变动情况的指数称为结构影响指数。

三个指数之间的数量关系为：

$$\frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_1}{\sum f_1}} \times \frac{\frac{\sum x_0 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}} = \frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}} \tag{7-14}$$

即：固定结构指数×结构影响指数=可变构成指数

这三个指数构成总平均指标两因素分析的指数体系。

由此可见，从动态的角度上看，总体平均水平的变动，既取决于各组水平的变动，也取决于总体内部各组结构的变动。

➪【例7-8】

我们以表7-7为例，说明总平均水平的变动中受总体各组水平变动和总体内各组结构变动的影响方向和程度。

表7-7　某企业工人工资及人数资料

工人类别	工资水平 x（元）		工人人数 f（人）	
	基期 x_0	报告期 x_1	基期 f_0	报告期 f_1
技术工人	600	700	600	800
辅助工人	400	500	400	700
合　　计	—	—	1 000	1 500

根据以上资料:

（1）计算可变构成指数

$$\frac{\dfrac{\sum x_1 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_0}{\sum f_0}}=\frac{\dfrac{700\times 800+500\times 700}{1\,500}}{\dfrac{600\times 600+400\times 400}{1\,000}}=\frac{606.67}{520}=1.166\,7=116.67\%$$

绝对数:

$$\frac{\sum x_1 f_1}{\sum f_1}-\frac{\sum x_0 f_0}{\sum f_0}=606.67-520=86.67\text{（元）}$$

表明：该企业工人总平均工资报告期比基期增长了 16.67%，平均每人增加 86.67 元。

（2）计算固定构成指数

$$\frac{\dfrac{\sum x_1 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_1}{\sum f_1}}=\frac{\dfrac{700\times 800+500\times 700}{1\,500}}{\dfrac{600\times 800+400\times 700}{1\,500}}=\frac{606.67}{506.67}=1.197\,4=119.74\%$$

绝对数:

$$\frac{\sum x_1 f_1}{\sum f_1}-\frac{\sum x_0 f_1}{\sum f_1}=606.67-506.67=100\text{（元）}$$

表明：由于该企业各类工人工资水平变动，使该企业工人总平均工资报告期比基期增长了 19.74%，平均每人增加了 100 元。

（3）计算结构影响指数

$$\frac{\dfrac{\sum x_0 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_0}{\sum f_0}}=\frac{\dfrac{600\times 800+400\times 700}{1\,500}}{\dfrac{600\times 600+400\times 400}{1\,000}}=\frac{506.67}{520}=0.974\,4=97.44\%$$

绝对数:

$$\frac{\sum x_0 f_1}{\sum f_1}-\frac{\sum x_0 f_0}{\sum f_0}=506.67-520=-13.33\text{（元）}$$

表明：该企业由于各类工人结构（人数）变动，使得该企业工人总平均工资下降了 2.56%，平均每人减少了 13.33 元。

（4）利用指数体系分析三者之间的数量关系

$$\frac{\dfrac{\sum x_1 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_1}{\sum f_1}}\times\frac{\dfrac{\sum x_0 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_0}{\sum f_0}}=\frac{\dfrac{\sum x_1 f_1}{\sum f_1}}{\dfrac{\sum x_0 f_0}{\sum f_0}}$$

即：$119.74\% \times 97.44\% = 116.67\%$

绝对数：$\left(\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1}\right) + \left(\frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0}\right) = \left(\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0}\right)$

即：$100 + (-13.33) = 86.67$（元）

整个计算分析表明：该企业工人总平均工资报告期比基期增长了 16.67%，平均每人增加 86.67 元，是由于各类工人工资水平变动，使工人总平均工资增长 19.74%，平均每人增加 100 元；又由于各类工人结构（人数）变动，使工人总平均工资下降 2.56%，平均每人减少 13.33 元的共同结果。

7.4 指数数列

7.4.1 指数数列的概念及种类

1．指数数列的概念

在统计的实践和理论研究中，为了反映客观事物在一个较长的时期不断发展变化的趋势，需要连续不断地编制指数，这便提出了指数数列的概念。所谓指数数列，就是将若干个同类现象的指数，按时间先后顺序排列而形成的数列。可见指数数列是一种时间数列，可用于复杂经济现象总体的动态分析。

2．指数数列的种类

指数数列按基期的选择不同，分为环比指数数列和定基指数数列；按同度量因素所固定的时期不同，分为可变权数指数数列和不变权数指数数列。

7.4.2 环比指数数列和定基指数数列

1．环比指数数列

在指数数列中，各期指数都以其前一期为基期来编制，称为环比指数数列。

（1）数量指标环比指数数列

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}, \frac{\sum q_2 p_1}{\sum q_1 p_1}, \frac{\sum q_3 p_2}{\sum q_2 p_2}, \cdots, \frac{\sum q_n p_{n-1}}{\sum q_{n-1} p_{n-1}}$$

（2）质量指标环比指数数列

$$\frac{\sum q_1 p_1}{\sum q_1 p_0}, \frac{\sum q_2 p_2}{\sum q_2 p_1}, \frac{\sum q_3 p_3}{\sum q_3 p_2}, \cdots, \frac{\sum q_n p_n}{\sum q_n p_{n-1}}$$

2．定基指数数列

在指数数列中，各期指数都以某一固定时期为基期来编制，称为定基指数数列。

（1）数量指标定基指数数列

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}, \frac{\sum q_2 p_0}{\sum q_0 p_0}, \frac{\sum q_3 p_0}{\sum q_0 p_0}, \cdots, \frac{\sum q_n p_0}{\sum q_0 p_0}$$

（2）质量指标定基指数数列

$$\frac{\sum q_1 p_1}{\sum q_1 p_0}, \frac{\sum q_2 p_2}{\sum q_2 p_0}, \frac{\sum q_3 p_3}{\sum q_3 p_0}, \cdots, \frac{\sum q_n p_n}{\sum q_n p_0}$$

环比指数数列和定基指数数列各有不同用途。若要说明各时期的现象与其前一时期对比变动的情况时，就可采用环比指数数列加以分析；而要说明各时期的现象与某一固定时期对比变动情况时，就应采用定基指数数列加以分析。

7.4.3　可变指数数列和不变指数数列

在前面介绍的指数编制过程中，我们知道指数中的同度量因素不仅起到同度量的作用，同时也起到了权数的作用，所以同度量因素也称权数。在指数数列中，若各时期指数的权数是不同的，则称为可变权数指数数列；若各时期指数的权数是相同的，则称为不变权数指数数列。

根据指数编制的一般原则，数量指标指数的同度量因素应固定在基期，质量指标指数的同度量因素应固定在报告期。因此在编制数量指标环比指数数列时，要求依次以前期为基期，这样同度量因素所属时期随基期变动而变动，故属于可变权数指数数列。在编制数量指标定基指数数列时，同度量因素固定在被比较的基期水平上，属于不变权数指数数列。而在编制质量指标指数数列时，由于要求同度量因素固定在报告期，所以不论是环比指数数列，还是定基指数数列，同度量因素所属时期总是随报告期变动而变动，故都属于可变权数指数数列。

在我国的统计实践中，工业产品数量的变动情况，是通过采用不变价格计算的工业总产值的变动来反映的。而采用的是不变权数来编制的工业产品产量指数数列。

设不变价格用 p_n 表示，则：

（1）环比产量指数数列

$$\frac{\sum q_1 p_n}{\sum q_0 p_n}, \frac{\sum q_2 p_n}{\sum q_1 p_n}, \frac{\sum q_3 p_n}{\sum q_2 p_n}, \frac{\sum q_4 p_n}{\sum q_3 p_n}\mathrm{L}$$

（2）定基产量指数数列

$$\frac{\sum q_1 p_n}{\sum q_0 p_n}, \frac{\sum q_2 p_n}{\sum q_0 p_n}, \frac{\sum q_3 p_n}{\sum q_0 p_n}, \frac{\sum q_4 p_n}{\sum q_0 p_n} \mathrm{L}$$

➪【例 7-9】

某市 1995—2000 年工业产品产量以 1990 年不变价格计算的工业总产值指数数列，如表 7-8 所示。

表 7-8　某市 1995—2000 年工业总产值指数数列

年　份	以 1990 年不变价格计算的工业总产值（亿元）	工业总产值指数（%）	
		环　比	定　基
1995	125	100	100
1996	136	108.8	108.8
1997	150	110.3	120.0
1998	166	110.7	132.8
1999	181	109.0	144.8
2000	198	109.4	158.4

这样可以观察该市 1995—2000 年期间工业生产增长变动情况。

从理论上讲，用不变价格编制的产值指数属不变权数指数，但是，在实际的统计工作中，编制的产值指数所采用的不变价格并不是永久不变的。因为由于时间久远，经济发展、产品品种与质量的不断变化，如果采用永久的不变价格，将会影响权数编制的代表性和正确性。

新中国成立以来，曾采用过 6 次不变价格（1950 年、1952 年、1957 年、1970 年、1980 年、1990 年）。当编制采用不是同一个不变价格的较长时期的总产值指数数列时，就要采用价格换算系数的方法来消除价格变动的影响。其公式如下：

$$\text{价格换算系数} = \frac{\text{交替年按不变价格计算的工业总产值}}{\text{交替年按旧不变价格计算的工业总产值}} \qquad (7\text{-}15)$$

➪【例 7-10】

1991 年为 20 世纪 80 年代不变价格过渡到 90 年代不变价格的交替年。上例中，该市 1991 年按 80 年代不变价格计算的工业总产值为 86 亿元，按 90 年代不变价格计算的工业总产值为 94 亿元，则

$$\text{价格换算系数}=\frac{\sum q_{1991}p_{90}}{\sum q_{1991}p_{80}}=\frac{94}{86}=1.093$$

利用价格换算系数就可进行按两种不变价格计算工业总产值对比。

上例中，该市 1985 年按 80 年代不变价格计算的工业总产值为 62 亿元，现在要求计算该市 2000 年与 1985 年的工业总产值指数，即计算：

$$\frac{\sum q_{2000}p_{90}}{\sum q_{1985}p_{80}}$$

可用两种计算方法：

（1）换算分母　　$\sum q_{1985}p_{80}\Rightarrow\sum q_{1985}p_{90}$

则　　$$\sum q_{1985}p_{90}=\sum q_{1985}p_{80}\times\frac{\sum q_{1991}p_{90}}{\sum q_{1991}p_{80}}=62\times1.093=67.77\text{（亿元）}$$

$$\frac{\sum q_{2000}p_{90}}{\sum q_{1985}p_{80}}\Rightarrow\frac{\sum q_{2000}p_{90}}{\sum q_{1985}p_{90}}=\frac{198}{67.77}=2.9217=292.17\%$$

也可将各年产值直接代入计算公式：

$$\frac{\sum q_{2000}\,p_{90}}{\sum q_{1985}\,p_{90}}=\frac{\sum q_{2000}\,p_{90}}{\sum q_{1985}\,p_{80}\times\dfrac{\sum q_{1991}\,p_{90}}{\sum q_{1991}\,p_{80}}}=\frac{\sum q_{2000}\,p_{90}}{\sum q_{1985}\,p_{80}}\times\frac{\sum q_{1991}\,p_{80}}{\sum q_{1985}\,p_{90}}$$

$$=\frac{198}{62}\times\frac{86}{94}=2.9217=292.17\%$$

（2）换算分子 $\sum q_{2000}p_{90}\Rightarrow\sum q_{2000}p_{80}$

则　　$$\sum q_{2000}p_{80}=\sum q_{2000}p_{90}\div\frac{\sum q_{1991}p_{90}}{\sum q_{1991}p_{80}}=198\div1.093=181.15\text{（亿元）}$$

$$\frac{\sum q_{2000}p_{90}}{\sum q_{1985}p_{80}}\Rightarrow\frac{\sum q_{2000}p_{90}}{\sum q_{1985}p_{80}}=\frac{181.5}{62}=2.9217=292.17\%$$

也可将各年总产值直接代入计算公式：

$$\frac{\sum q_{1991}p_{80}}{\sum q_{1985}p_{80}}=\frac{\sum q_{2000}p_{90}\div\dfrac{\sum q_{1991}p_{90}}{\sum q_{1991}p_{80}}}{\sum q_{1985}p_{80}}=\frac{\sum q_{2000}p_{90}}{\sum q_{1985}p_{80}}\times\frac{\sum q_{1991}p_{80}}{\sum q_{1991}p_{90}}$$

$$=\frac{198}{62}\times\frac{86}{94}=2.9217=292.17\%$$

两种方法的计算结果一致，均表明该市工业生产的变动，2000 年比 1985 年增长了 192.17%。

本章小结

1. 从广义而言，凡是反映社会经济现象数量差异和变动程度的相对数，都可称为指数。狭义的或严格意义的指数，是指测定复杂的、不可同度量的社会现象综合变动的相对数。指数在社会经济统计中除用于测定诸如价格、产量等综合变动外，还用于总量变动的因素分析和对多指标复杂事物的综合测评。

2. 指数按研究范围不同分为个体指数和总指数，以及介于其中的类指数；按指标的性质不同分为数量指标指数和质量指标指数。总指数有综合指数和平均数指数两种形式。

在资料完全相同的情况下，平均数指数可视为综合指数的变形，计算结果相同。它们的主要区别是：综合指数通常用全面资料计算；先综合，后对比；公式中分子分母之差可以说明变动因素影响的差额。平均数指数用抽样资料计算；先对比，后综合；分子分母之差不能说明变动因素影响的总量差额。在编制方法上，综合指数的关键是确定同度量因素；平均数指数的关键是确定权数。

3. 总平均数指数属于广义指数范畴，它根据分组资料计算，有可变构成指数、固定构成指数、结构变动影响指数三种形式，在结构分析和因素分析中有独特的意义。

4. 依据互相联系的一组指数以数量平衡形式构成的指数体系，可以对总量变动进行因素分析。

5. 居民消费价格指数和股价指数等是几种最常见的重要的经济指数。从编制方法中，可以看到编制指数在资料采集、公式选择、权数和基期的确定等方面需要处理一系列技术性问题；从指数的应用中可以看到指数在社会经济问题研究中的重要意义。

复习思考题

一、概念题

广义指数　狭义指数　个体指数　总指数　数量指数　质量指数　综合指数　加权算术平均指数　加权调和平均指数　指数体系　指数因素分析

二、简答题

1. 什么是统计指数？编制统计指数有何重要作用？
2. 统计指数有哪些种类？
3. 什么是数量指标指数和质量指标指数？编制其指数时应遵循的一般原则是什么？
4. 什么是同度量因素？它在指数编制中有什么作用？
5. 什么是平均数指数？它的应用条件和编制原则是什么？
6. 简要说明我国零售物价总指数的编制方法。
7. 什么是指数体系？它有什么作用？

8. 什么是因素分析？因素分析包括哪两方面内容？
9. 什么是不变构成指数、固定构成指数和结构影响指数？
10. 编制指数数列有什么重要定义？什么是环比指数数列和定基指数数列？什么是可变权数数列和不变权数数列？

三、练习题

1．判断题（把“√”或“×”填在题后的括号里）

（1）广义指数就是指各种相对数。（　　）

（2）劳动生产率指数和单位成本指数都是数量指标指数。（　　）

（3）为使价格指数的计算符合现实经济意义，在编制价格总指数时，同度量因素应是基期的商品销售量。（　　）

（4）平均指数这种编制总指数的方法，既可用于全面调查的资料，又可用于非全面调查的资料。（　　）

（5）数量指标指数可以变形为加权调和平均指数形式。（　　）

（6）统计指数是综合反映社会经济现象总变动方向及变动幅度的相对数。（　　）

（7）固定构成指数=可变构成指数×结构影响指数。（　　）

（8）利用派氏指数公式计算总指数的优点在于考虑到现实的经济意义。（　　）

2．单选题

（1）狭义指数是指（　　）。

A. 反映价格变动的相对数
B. 反映动态变动的相对数
C. 反映物量变动的相对数
D. 反映不能直接相加和对比的复杂现象综合变动的相对数

（2）用综合指数法编制总指数的关键问题之一是（　　）。

A. 确定被比对象　　B. 确定同度量因素及其固定时期
C. 确定对比基期　　D. 计算个体指数

（3）统计指数划分为个体指数和总指数的依据，是按指数（　　）。

A. 反映的对象范围不同　　B. 同度量因素不同
C. 计算时是否进行加权　　D. 指数化的指标不相同

（4）以 p 代表质量指标，q 代表数量指标，下标 0 表示基期，1 表示报告期，则质量指标综合指数公式是（　　）。

A. $\dfrac{\sum p_1q_1}{\sum p_0q_0}$　　B. $\dfrac{\sum p_0q_1}{\sum p_1q_1}$　　C. $\dfrac{\sum p_1q_1}{\sum p_0q_1}$　　D. $\dfrac{\sum p_1q_0}{\sum p_1q_1}$

（5）下列指数中，（　　）是商品销售额综合指数。

A. $\frac{\sum p_1q_1}{\sum p_0q_0}$　　B. $\frac{\sum p_1q_0}{\sum p_1q_1}$　　C. $\frac{\sum p_0q_1}{\sum p_0q_0}$　　D. $\frac{\sum p_0q_1}{\sum p_1q_1}$

（6）某商店商品销售额报告期和基期相同，商品价格报告期比基期提高了5%，那么报告期商品销售量比基期（　　）。

A. 提高了10%　B. 减少了4.8%　C. 增长了5.2%　D. 减少了10%

（7）若将加权算术平均指数变形为综合指数，其分母应为（　　）。

A. $p_0\ q_0$　　B. $p_1\ q_1$　　C. p_1q_0　　D. p_0q_1

（8）为了反映两个同类企业劳动生产率的提高情况，则需要编制（　　）。

A. 数量指标综合指数　　B. 可变构成指数

C. 结构影响指数　　D. 固定构成指数

3．多选题

（1）同度量因素的作用有（　　）。

A. 平均作用　　B. 权数作用　　C. 比较作用　　D. 稳定作用

E. 同度量作用

（2）编制综合指数的一般原则是（　　）。

A. 质量指标指数以报告期数量指标作为同度量因素

B. 数量指标指数以报告期质量指标作为同度量因素

C. 质量指标指数以基期数量指标作为同度量因素

D. 数量指标指数以基期质量指标作为同度量因素

E. 随便确定

（3）若用 p 表示价格，q 表示商品销售量，则公式 $\sum p_1q_1-\sum p_0q_1$ 的意义是（　　）。

A. 综合反映价格变动和销售量变动的绝对额

B. 综合反映多种商品价格变动而增减的销售额

C. 综合反映由于价格变动而使消费者增减的货币支出额

D. 综合反映多种商品销售量变动而增减的销售额

E. 综合反映商品销售额本身变动而增减的绝对额

（4）下列关于综合指数的表述正确的是（　　）。

A. 综合反映多种现象的平均变动程度

B. 两个总量指标对比的动态相对数

C. 固定一个或一个以上因素，反映另一个因素的变动

D. 分子与分母是两个或两个以上因素乘积之和

E. 分子或分母中有一项假定指标

（5）下列指数中，属于质量指标指数的有（　　）。

A. 农产品产量总指数　　B. 农产品收购价格总指数

C. 某种工业产品成本总指数　　D. 全部商品批发价格指数

E. 职工工资个体指数

（6）编制总指数时，必须注意的是（　　）。

A. 综合指数法一般使用全面资料

B. 平均指数法可以使用非全面资料

C. 平均指数法等同于综合指数法

D. 平均指数法在一定条件下可以是综合指数法变形

E. 综合指数法是平均指数法的变形

（7）下面能反映平均指标变动的指数是（　　）。

A. 可变构成指数　　B. 固定构成指数　　C. 结构影响指数

D. 算术平均指数　　E. 调和平均指数

（8）若用某企业职工人数和劳动生产率的分组资料来进行分析，则该企业总的劳动生产率的变动主要受到（　　）。

A. 企业全部职工人数变动的影响

B. 企业劳动生产率变动的影响

C. 企业各类职工人数占全部职工人数比重的变动影响

D. 企业各类工人劳动生产率的变动影响

E. 各组职工人数和相应劳动生产率两因素的影响

4．计算题

（1）某商业企业经营三种商品，其基期和报告期的有关资料如表 7-9 所示。试分析销售量综合变动和价格综合变动及其对商品销售总额的影响。

（2）某企业生产资料如表 7-10 所示。试计算：

1）三种产品的产量总指数及由于产量提高而增加的总成本；

2）单位产品成本总指数及由于单位产品成本变动而影响的总成本。

表 7-9　某企业三种商品的基期和报告期资料

商品名称	计量单位	销售量		价格（元）	
		基　期	报告期	基　期	报告期
甲	件	100	125	1 500	1 500
乙	吨	50	50	3 000	1 150
丙	米	300	270	980	1 200

表 7-10　某企业生产资料

产品名称	总成本（万元）		产量增长率（%）
	报告期	基　期	
甲	240	200	25
乙	485	450	10
丙	480	350	40
合　计	1 205	1 000	—

（3）某企业销售资料如表 7-11 所示。试计算：

1）两种产品销售价格平均降低程度及由于价格降低而减少的销售额；

2）两种产品销售量总指数及由于产品销售量变动而增减的销售额。

表 7-11　某企业销售资料

产品名称	销 售 额（万元）		价格降低率（%）
	基　期	报告期	
甲	750	780	5
乙	500	520	3
合　计	1 250	1 300	—

（4）某厂各种可比产品总成本，2004 年为 30 万元，比 2003 年增加 5 万元，单位产品成本 2004 年比 2003 年平均降低 3%。试计算：

1）产品总成本指数和产品产量指数；

2）由于单位产品成本降低而节约的绝对额。

（5）某公司所属两个企业工人劳动生产率及工人人数资料如表 7-12 所示。试从相对数和绝对数两个方面：

1）反映全公司工人劳动生产率的变动情况；

2）分析由于各企业工人劳动生产率和各企业工人人数变动对全公司工人劳动生产率变动的影响程度。

表 7-12　甲、乙两个企业的劳动生产率及人数

企业名称	劳动生产率（元/人）		工人人数（人）	
	2007 年	2008 年	2007 年	2008 年
甲	11 000	18 000	800	700
乙	19 000	35 000	900	1 200
合 计	—	—	1 700	1 900



（6）某企业工人工资资料如表 7-13 所示。试计算：

1）说明该公司职工总平均工资变动程度及增（减）的绝对额；

2）总平均工资变动中各因素影响的程度及绝对额。

表 7-13　某企业工人工资资料

工人类别	月平均工资（元）		工人人数（人）	
	基　期	报告期	基　期	报告期
技术工人	1 140	1 350	100	100
辅助工人	960	1 200	120	150
合计	—	—	220	250

（7）某工业企业的资料如表 7-14 所示。试分析：原材料费用总额受产品产量、单位产品原材料消耗量和原材料价格变动影响程度。

表 7-14　某企业资料情况

产品名称	计量单位	产　量		单位原材料消耗量（公斤）		原材料价格（元/公斤）	
		基　期	报告期	基　期	报告期	基　期	报告期
甲	套	80	90	110	105	100	10.5
乙	件	100	120	25	28	15.2	10.8
丙	台	50	55	210	200	18.0	17.0

CHAPTER 8

第 8 章　Excel在统计分析中的运用

章首引例

目前，统计分析方法的计算机处理主要是使用统计分析软件包作为工具，统计分析工具有 SPSS、SAS、Eview、Excel 等软件。微软公司的 Excel 2000（以及更高级版本）提供了一些基本的统计分析工具，包括常用的统计函数和统计宏软件包。与 SPSS 等专业统计工具相比，Excel 更方便、更灵活，但在方法的深度与广度上无法与 SPSS 等专业统计软件相比。在一般企业管理中，Excel 提供的统计功能基本上可以满足大多数统计工作者的需要。在本章中，我们主要介绍 Excel 的统计分析功能在统计计算方面的应用。

本章学习目标

- ☑ 熟练掌握 Excel 数据分析工具的安装过程；
- ☑ 熟练掌握 Excel 实现统计数据整理以及数据描述；
- ☑ 学会使用 Excel 进行区间估计；
- ☑ 熟练掌握 Excel 实现数列的相关和回归分析，特别是时间数列趋势分析；
- ☑ 学会使用 Excel 进行指数分析。

8.1 Excel数据分析程序的安装

如果在“工具”菜单中没有“数据分析”命令，必须在 Microsoft Excel 中安装“分析工具库”。方法是：进入 Windows 2000 下的 Microsoft Excel（7.0 版本），单击“工具”菜单中的“加载宏”，在弹出“加载宏”对话框中勾选“分析工具库”，如图 8-1 所示，单击“确定”按钮，即完成了 Excel 数据分析程序的安装。在 Excel 的“工具”菜单里将出现“数据分析”的命令，如图 8-2 所示。

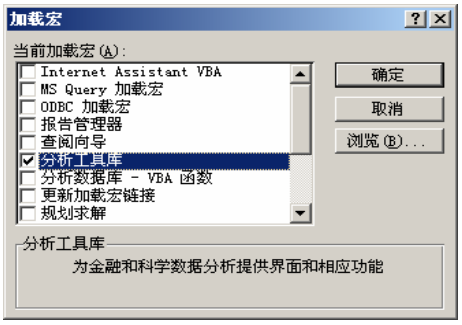


图 8-1　“加载宏”对话框

完成 Excel 数据分析程序安装后，单击“工具”菜单中的“数据分析”，即会弹出 Excel 的“数据分析”对话框，从中选择某一个统计分析工具，单击“确定”按钮，就能进入该统计分析工具的运行状态。

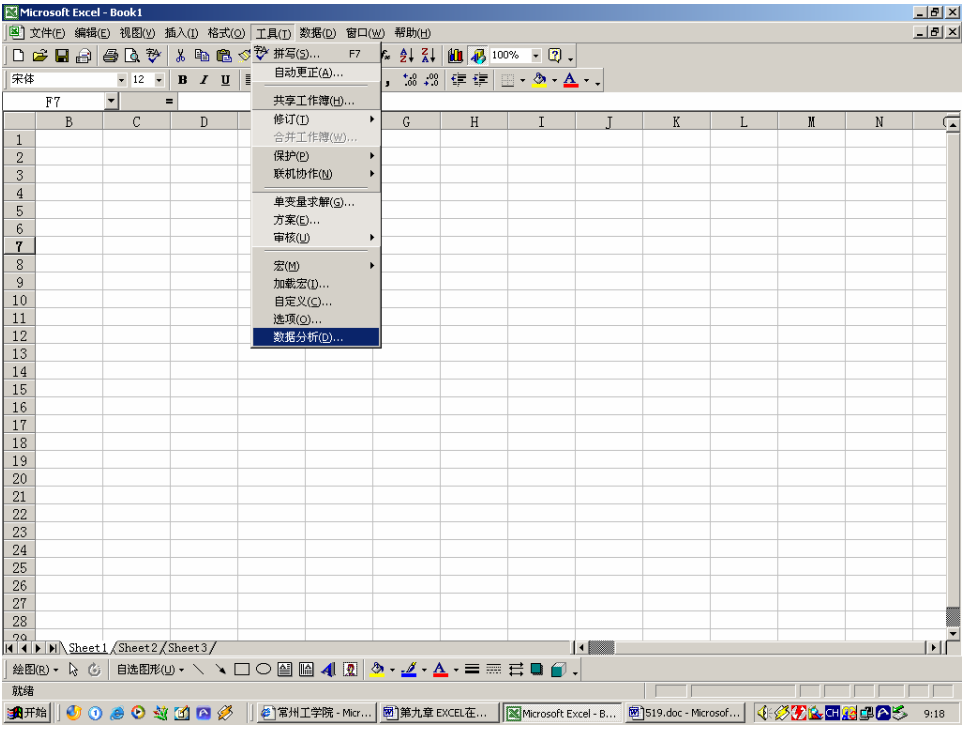


图 8-2　“工具”菜单的“数据分析”命令

如果现在想进行相关系数分析，请单击“工具”菜单中的“数据分析”，弹出如图 8-3 所示的对话框。

选择“相关系数”，单击“确定”按钮，即可在“相关系数”对话框中进行相关系数的分析，如图 8-4 所示。



图 8-3　“数据分析”对话框

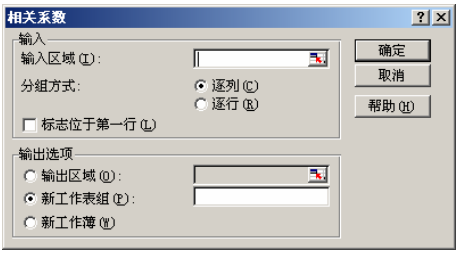


图 8-4　“相关系数”对话框

此外，Excel 数据分析还能结合统计函数“f_x”进行简单或复杂的编辑运算。单击任一单元格，选择工具栏上的 f_x 按钮，出现如图 8-5 所示的对话框，在其中可进行编辑。

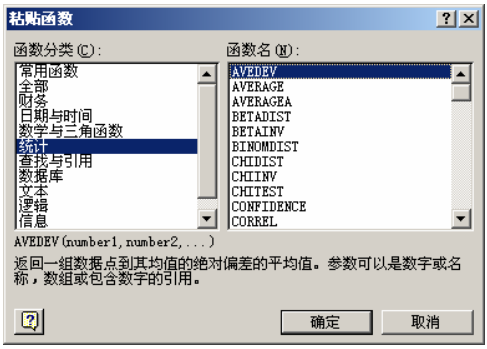


图 8-5 “粘贴函数”对话框

8.2 用Excel整理统计数据

8.2.1 统计分组

用 Excel 进行统计分组有两种方法，一是利用 FREQUENCY 函数；二是利用数据分析中的“直方图”工具。

对图 8-6 所示的数据进行统计分析。

1．用FREQUENCY函数编制频数表

第一步：确定每一组的上限值，确定上限值是编制频数表的关键，确定了上限值实际就确定了每一组的组距和组限。本例输入的上限值分别为 900，1 100，1 300，1 500，1 700，并且把这些上限值输入到 A7：A11。

第二步：选取结果存放的单元格区域 B7：B11。

第三步：在编辑栏输入公式“=frequency（A1：J5，A7：A11）”，按 Ctrl+Shift+Enter 组合键，即可获得各组相应的频数，结果如图 8-7 所示。

	A	B	C	D	E	F	G	H	I	J
1	780	830	860	870	880	930	1010	1010	1030	1050
2	1050	1070	1080	1080	1100	1100	1130	1140	1150	1160
3	1170	1170	1180	1180	1190	1200	1210	1230	1230	1250
4	1250	1250	1260	1260	1270	1270	1310	1320	1350	1360
5	1370	1380	1380	1410	1420	1420	1460	1510	1580	1680

图 8-6 已知数据

7	900	5
8	1100	11
9	1300	20
10	1500	11
11	1700	3

图 8-7 频数表

2．采用数据分析工具做频数分布表

第一步：在“工具”菜单中单击“数据分析”选项，从打开对话框的“分析工具”列表框中选择“直方图”，打开“直方图”对话框，如图 8-8 所示。

第二步：在“输入区域”输入“$A$1：$J$5”，在“接收区域”输入“$A$7：$A$11”。接收区域指的是分组标志所在的区域，假定我们把分组标志输入到 A7：A11 单元格，注意这里只能输入每一组的上限值，即 900，1 100，1 300，1 500，1 700。

第三步：选择输出选项，可选择“输出区域”、“新工作表组”或“新工作簿”。在这里选择“输出区域”，可以直接选择一个区域，也可以直接输入一个单元格（表示输出区域的左上角）。这里推荐只输入一个单元格（本例为 D10），因为事先不知道具体的输出区域有多大。

第四步：选择“图表输出”，可以得到直方图；选择“累计百分率”，系统将在直方图上添加累计频率折线；选择“柏拉图”，可得到按降序排列的直方图。

第五步：单击“确定”按钮，可得输出结果，如图 8-9 所示。

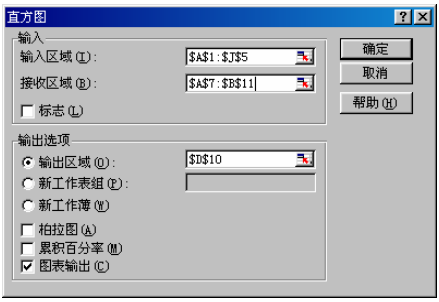


图 8-8　“直方图”对话框

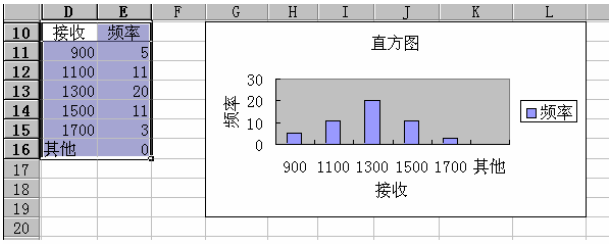


图 8-9　频数分布和直方图

注意：图 8-9 实际上是一个“草图”，还需要做如下修改：

① 单击任一直条，然后右击，在弹出的快捷菜单中选择“数据系列格式”，弹出“数据系列格式”对话框，从中选择“选项”标签，调整“间距宽度”为 0，如图 8-10 所示。

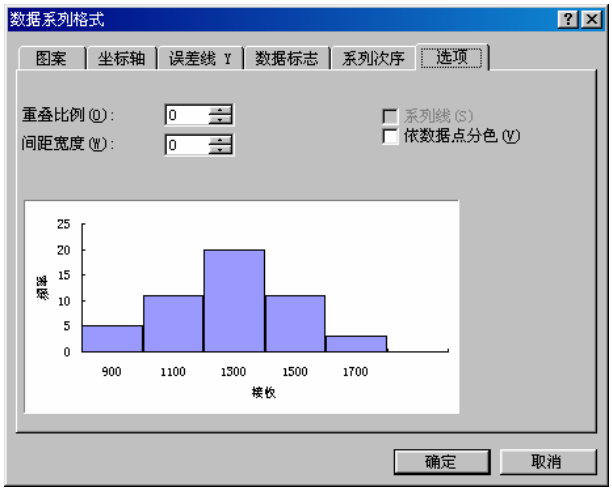


图 8-10　“数据系列格式”对话框

② 在图域的非直条处点右键，清除（灰色）背景色。

③ 将 D16 中的“其他”清除，直方图中的“其他”直条消失。

④ 在图例上右击清除图例“频率”，在图表标题上右击清除图表标题“直方图”。

⑤ 在绘图区域右击，从快捷菜单中选中“图表选项”，在“分类轴”下框内输入“工厂日产量直方图”，在“数值轴”下框内输入“频率”，如图 8-11 所示。

⑥ 调整后的直方图如图 8-12 所示。

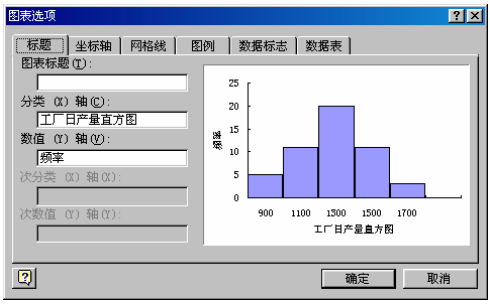


图 8-11 “图表选项”对话框

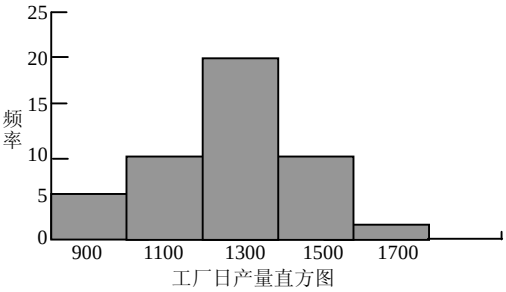


图 8-12 调整后的直方图

8.2.2 做统计图

Excel 提供的统计图有多种，包括柱形图、条形图、折线图、饼图、散点图、面积图、环形图、雷达图、曲面图、气泡图、股价图、圆柱图和圆锥图等，各种图的做法大同小异。表 8-1 是某城市居民关注广告类型的次数分布。

表 8-1 某城市居民关注广告类型

广告类型	人数（人）	比 例	频率（%）
商品广告	112	0.560	56.0
服务广告	51	0.255	25.5
金融广告	9	0.045	4.5
房地产广告	16	0.080	8.0
招生招聘广告	10	0.050	5.0
其他广告	2	0.010	1.0
合 计	200	1	100.0

1. 饼图

利用表 8-1 的数据绘制饼图，数据输入后如图 8-13 所示。

第一步：选中某一个单元格，单击“插入”菜单，选择“图表选项”，弹出“图表类型”

对话框，如图 8-14 所示。

第二步：在“图表类型”中选择“饼图”，然后在“子图表类型”中选择一种类型。这里我们选择“分离型三维饼图”，然后单击“下一步”按钮，打开“图表源数据”对话框，如图 8-15 所示。



	A	B	C	D	E
1	广告类型	人数（人）	比例	频率（%）	
2	商品广告	112	0.560	56.0	
3	服务广告	51	0.255	25.5	
4	金融广告	9	0.045	4.5	
5	房地产广告	16	0.080	8.0	
6	招生招聘广告	10	0.050	5.0	
7	其他广告	2	0.010	1.0	
8	合计	200	1	100.0	
9					

图 8-13　Excel 中的某城市居民关注广告类型的统计



图 8-14　“图表类型”对话框

第三步：在“图表源数据”对话框中输入数据所在区域，本例选择系列产生在“列”，单击“完成”按钮，即可得到如图 8-16 所示的饼图。



图 8-15　“图表源数据”对话框

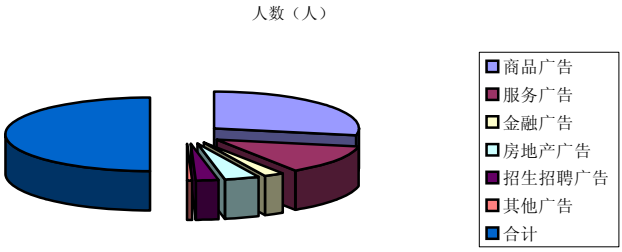


图 8-16　饼图

2．条形图

做条形图的步骤如下：

第一步：选中某一个单元格，单击“插入”菜单，选择“图表选项”，弹出“图表类型”

对话框，如图 8-17 所示。

第二步：在“图表类型”中选择“条形图”，然后在“子图表类型”中选择一种类型，这里我们选择第五种。然后单击“下一步”按钮，打开“图表源数据”对话框，如图 8-18 所示。



图 8-17 “图表类型”对话框



图 8-18 “图表源数据”对话框

第三步：在“图表源数据”对话框中填入数据所在区域，本例为选择系列产生在“列”，单击“完成”按钮，即可得到如图 8-19 所示的条形图。

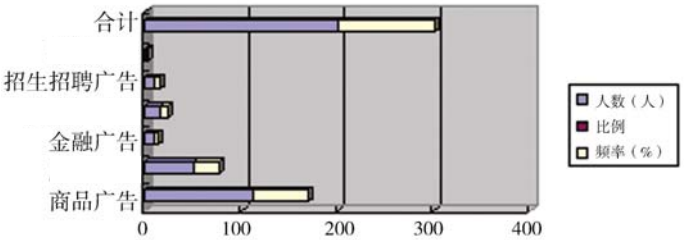


图 8-19 条形图

8.3 用Excel计算描述统计量

8.3.1 用函数计算描述统计量

常用的描述统计量有众数、中位数、算术平均数、调和平均数、标准差、标准差系数等。一般来说，在 Excel 中求这些统计量，未分组资料可用函数计算，已分组资料可用公式计算。

1．众数

某城市居民家庭收入情况如表 8-2 所示，求众数。

计算步骤如下：

第一步：创建 Excel 文件，输入表 8-2 中的数据。

第二步：在 B11 单元格中输入公式“=20000+（（260–240）/（（260–240）+（260–140））*5000”，结果如图 8-20 所示。

表 8-2　某城市居民家庭收入

年收入水平（元）	居民户数 f
10 000 以下	92
10 000～15 000	180
15 000～20 000	240
20 000～25 000	260
25 000～30 000	140
30 000～35 000	53
35 000 以上	35
合　　计	1 000

2．中位数

下面仍以表 8-2 中的数据，计算中位数。计算结果如图 8-21 所示。

计算步骤如下：

第一步：确定中位数组，即第三组。

第二步：利用上限公式，在 B11 单元格中输入公式“=20000–（500–488）/240*5000”。

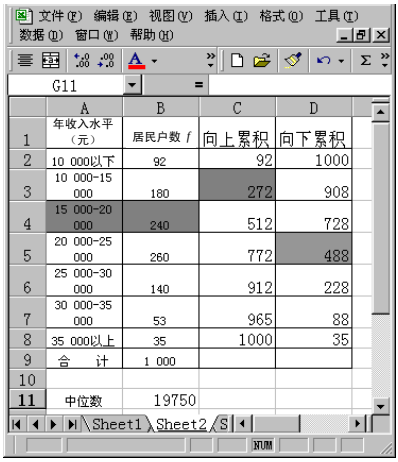


图 8-20　用 Excel 计算众数

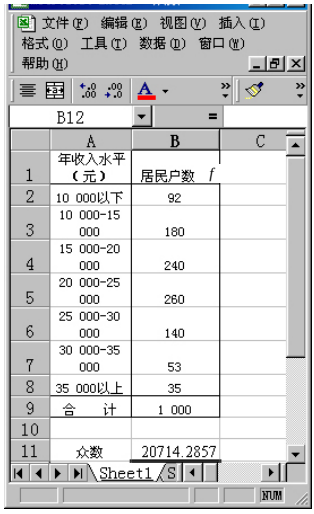


图 8-21　用 Excel 计算中位数

3．算术平均数

算术平均数是最常用、最基本的一种平均指标。

表 8-3 是某公司 50 名员工的月工资情况。试求该公司员工月工资的一般水平。计算结果如图 8-22 所示。

表 8-3　某公司员工月工资基本情况

工资 x（元）	员工数 f（人）
800	5
1 000	10
1 200	20
1 500	7
2 000	5
2 500	3
合计	50

计算步骤如下：



	A	B	C
1	工资 x（元）	员工数 f（人）	xf
2	800	5	4000
3	1000	10	10000
4	1200	20	24000
5	1500	7	10500
6	2000	5	10000
7	2500	3	7500
8	合计	50	66000
9			
10	工资平均数		1320

图 8-22　用 Excel 计算算术平均值

第一步：创建 Excel 文件，输入表 8-3 的数据。

第二步：在 C2 单元格中输入公式“=A2*B2”，拖拽鼠标将公式复制到 C3：C7。

第三步：在 C8 单元格中输入公式“=SUM(C2：C7)”。

第四步：在 C10 单元格中输入公式“=C8/B8”，即求得算术平均数。

4．调和平均数

调和平均数是根据标志值的倒数计算的，它是标志值倒数的算术平均数的倒数。

某工厂本月购进某种材料四批，每批价格以及金额如表 8-4 所示，求该材料的平均价格。计算结果如图 8-23 所示。

表 8-4　某工厂购物情况

	价格 x（元/千克）	采购金额 m（元）
第一批	35	10 000
第二批	40	20 000
第三批	45	15 000
第四批	50	5 000
合　计	—	50 000

计算步骤如下：

第一步：将数据输入 Excel 文件。

第二步：在 D2 单元格输入公式“= C2/B2”，拖拽鼠标将公式复制到 D3:D5。

第三步：在 D6 单元格输入公式“= SUM（D2:D5）”。

第四步：在 B8 单元格输入公式“= C6/D6”，即求得平均价格。

5．几何平均数

某笔为期 20 年的投资按复利计算收益，前 10 年的年利率为 10%，中间 5 年的年利率为 8%，最后 5 年的年利率为 6%。求整个投资期间的年平均利率。

计算步骤如下：

第一步：在 Excel 中，输入每一年的本利。

第二步：在 D2 单元格输入“= GEOMEAN（B2：B21）–1”，即可求得年均利率。计算结果如图 8-24 所示。

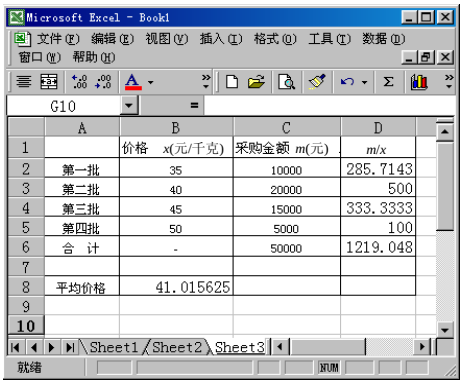


图 8-23　用 Excel 计算调和平均值

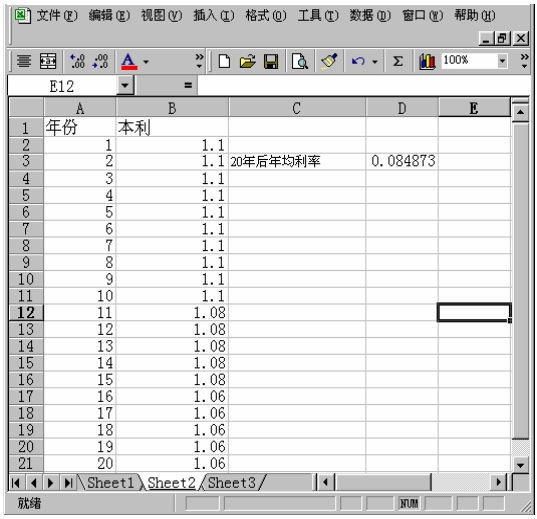


图 8-24　用 Excel 计算几何平均数

6．标准差

标准差是统计中最常用、最重要的变异指标，其意义与平均差相同，适用于定距数据和定比数据。

根据表 8-3 的数据求员工月工资标准差，计算结果如图 8-25 所示。

计算步骤如下。

第一步：利用图 8-22 中的数据，在 D2 单元格输入公式 “=A2–$C$10”。注意 C10 一定要用绝对引用。

第二步：在 E2 单元格中输入公式“=ABS（D2）*B2”，拖拽鼠标将公式复制到 E3：E7。

第三步：在 F2 单元格中输入公式“=D2*D2*B2”，拖拽鼠标将公式复制到 F3：F7。

第四步：在 F8 单元格中输入公式“=SUM（F2：F7）”。

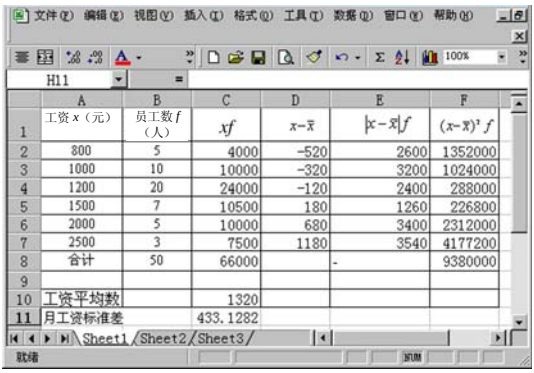
第五步：在 C10 单元格中输入公式“=SQRT（F8/B8）”，即求得标准差。

7．标准差系数

标准差系数是数据的标准差与其对应的均值之比。

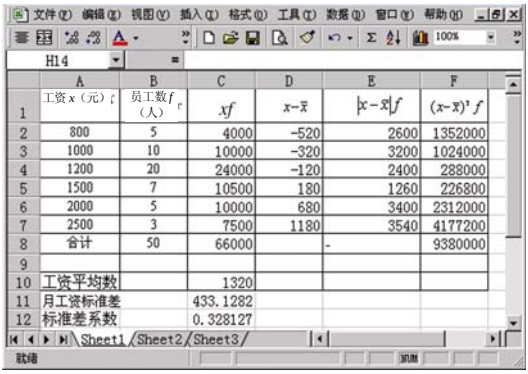
根据表 8-3 的数据求员工月工资标准差系数，计算结果如图 8-26 所示。

计算步骤是：在标准差的计算结果上，在 C12 单元格中输入公式“=C11/C10”即可。



	A	B	C	D	E	F
1	工资 x（元）	员工数 f（人）	xf	$x-\bar{x}$	$\lvert x-\bar{x}\rvert f$	$(x-\bar{x})^2 f$
2	800	5	4000	-520	2600	1352000
3	1000	10	10000	-320	3200	1024000
4	1200	20	24000	-120	2400	288000
5	1500	7	10500	180	1260	226800
6	2000	5	10000	680	3400	2312000
7	2500	3	7500	1180	3540	4177200
8	合计	50	66000		-	9380000
9						
10	工资平均数		1320			
11	月工资标准差		433.1282			

图 8-25　用 Excel 计算标准差



	A	B	C	D	E	F
1	工资 x（元）	员工数 f（人）	xf	$x-\bar{x}$	$\lvert x-\bar{x}\rvert f$	$(x-\bar{x})^2 f$
2	800	5	4000	-520	2600	1352000
3	1000	10	10000	-320	3200	1024000
4	1200	20	24000	-120	2400	288000
5	1500	7	10500	180	1260	226800
6	2000	5	10000	680	3400	2312000
7	2500	3	7500	1180	3540	4177200
8	合计	50	66000		-	9380000
9						
10	工资平均数		1320			
11	月工资标准差		433.1282			
12	标准差系数		0.328127			

图 8-26　用 Excel 计算标准差系数

8.3.2　用数据分析工具描述统计量

根据表 8-3 中的数据，利用数据分析工具描述统计量。

首先把数据输入到单元格，按下面步骤操作：

第一步：在“工具”菜单中选择“数据分析”选项，从其对话框中选择“描述统计”，单击“确定”按钮后打开“描述统计”对话框，如图 8-27 所示。

第二步：在“输入区域中”输入“$A$1：$A$51”，在“输出区域”中输入“$C$1”，其他复选框可根据需要选定，选择“汇总统计”，可给出一系列描述统计量。

第三步：单击“确定”按钮，即可得到如图 8-28 所示的结果。

8.4　用Excel进行区间估计

用 Excel 求区间估计通常需要根据区间估计的要求逐步进行。我们回忆一下，进行区间估计首先要计算样本均值，即点估计值。该值是要估计区间的中心；接着计算样本标准差，在此基础上结合样本量构造抽样误差，再结合置信度构造极限误差，样本均值分别加

上、减去极限误差即得到区间估计上下限。下面结合一个例子具体说明操作步骤。



图 8-27　“描述统计”对话框

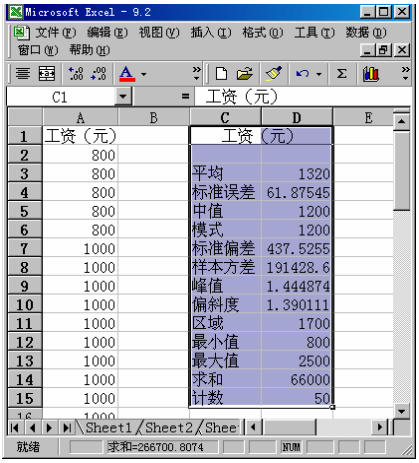


图 8-28　描述统计输出结果

某加工企业生产一种零件，对某天加工的零件每隔一段时间进行抽样，共取出 10 个，测得其重量（e）如下：

样本数据：11.10，10.90，11.00，10.94，10.97，10.91，11.21，11.16，11.04，l0.98 假定零件重量服从正态分布，试以 95%的置信水平估计该企业生产的零件平均重量的置信区间，结果如图 8-29 所示。

计算步骤如下。

第一步：把数据输入到 A2：A11 单元格。

第二步：在 C2 单元格输入公式“=COUNT(A2：A11)”。

第三步：在 C3 单元格输入公式“=AVERAGE（A2：A11)”。

第四步：在 C4 单元格输入公式“=STDEV(A2：A11)”。

第五步：在 C5 单元格输入公式“=C4/SQRT（C2)”。

第六步：在 C6 单元格输入数据“0.95”。

第七步：在 C7 单元格输入公式“=C2–1”。

第八步：在 C8 单元格输入公式“=TINV(1–C6，C7)”。

第九步：在 C9 单元格输入公式“=C8*C5”。

第十步：在 C10 单元格输入公式“=C3–C9”。

第十一步：在 C11 单元格输入公式“=C3+C9”。

所以，置信区间是[10.95，11.10]。



图 8-29　区间估计结果

8.5 用Excel进行相关分析与回归分析

现在以某地区的8个同类工业企业2004年的有关月产量与生产费用的资料为例进行相关分析和回归分析。

8.5.1 相关分析

1．散点图

散点图也称相关图，它是把相关表中两个变量的对应值在平面直角坐标系中用点描绘出来，表示相关点的分布情况。

根据表 8-5 中的数据，画出散点图。

表 8-5　某地区 2004 年月产量与生产费用情况

企业编号	月产量 x（千吨）	生产费用 y（万元）
1	1.2	62
2	2.0	86
3	3.1	80
4	3.8	110
5	5.0	115
6	6.1	132
7	7.2	135
8	8.0	160

操作步骤如下。

第一步：将数据输入 Excel 文件。

第二步：单击“插入”菜单，选择“图表”，或者单击工具栏中的图表向导按钮，弹出“图表类型”对话框。在“标准类型”中选择“XY 散点图”，如图 8-30 所示。

第三步：单击“下一步”按钮，选择“数据区域”，本题为$B$1：$C$9，如图 8-31 所示。

第四步：单击“下一步”按钮，将 X 轴、Y 轴的名称写上，如图 8-32 所示。

第五步：单击“下一步”按钮，选择“作为其中的对象插入”，如图 8-33 所示。

单击“完成”按钮，最后结果如图 8-34 所示。

2．相关系数的计算

用 Excel 进行相关分析有两种方法，一是利用相关系数函数，另一种是利用相关分析宏。

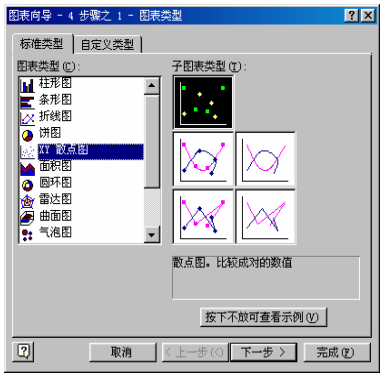


图 8-30　“图表类型”对话框

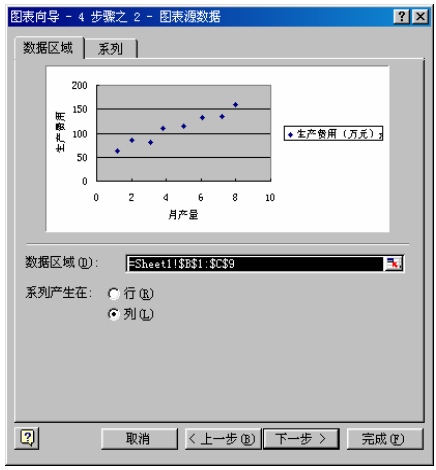


图 8-31　“图表源数据”对话框

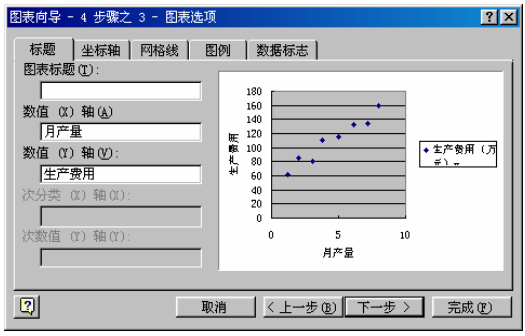


图 8-32　“图表选项”对话框



图 8-33　“图表位置”对话框

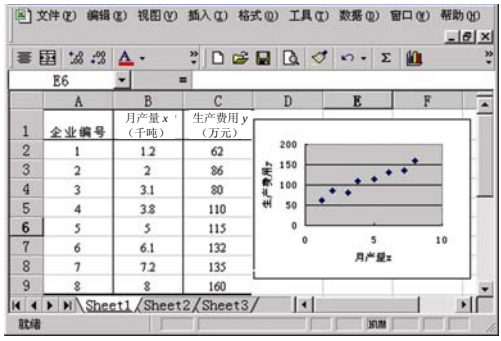


图 8-34　散点图

（1）利用相关系数函数

在 Excel 中，提供了两个计算两个变量之间相关系数的方法，CORREL 函数和 PERSON

函数，这两个函数是等价的，这里我们介绍 CORREL 函数计算相关系数。

第一步：单击任意一个空白单元格，单击“插入”菜单，选择“函数”选项，打开“粘贴函数”对话框。在“函数分类”中选择“统计”，在“函数名”中选择 CORREL，如图 8-35 所示。单击“确定”后出现 CORREL 对话框。

第二步：在 Array1 中输入“B2：B9”，在 Array2 中输入“C2：C9”，即可在对话框下方显示出结果为 0.9697，结果如图 8-36 所示。

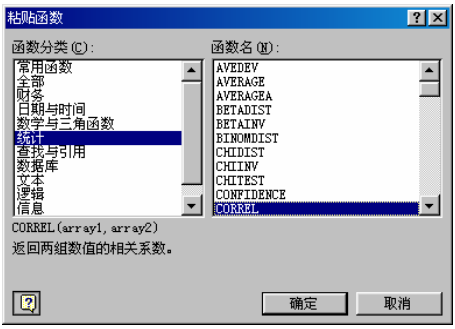


图 8-35 “粘贴函数”对话框



图 8-36 “CORREL”对话框

（2）用相关系数宏计算相关系数

第一步：单击“工具”菜单，选择“数据分析”选项，在“数据分析”选项中选择相关系数，弹出“相关系数”对话框，如图 8-37 所示。

第二步：在“输入区域”输入“$B$1：$C$9”，“分组方式”选择“逐列”，选择“标志位于第一行”，在“输出区域”中输入“$D$1”，单击“确定”，输出结果如图 8-38 所示。

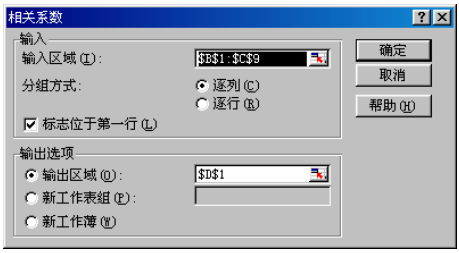


图 8-37 “相关系数”对话框



图 8-38 相关分析输出结果

在上面的输出结果中，月产量和生产费用的自相关系数均为 1，产量和生产费用的关系为 0.9697，和函数计算的完全相同。

8.5.2　一元线性回归分析

一元线性回归分析方法的具体步骤如下。

第一步：单击“工具”菜单，选择“数据分析”选项，出现“数据分析”对话框，在“分析工具”中选择“回归”，如图 8-39 所示。

第二步：单击“确定”按钮，弹出“回归”对话框。在“Y 值输入区域”输入“$C$1：$C$9”，在“X 值输入区域”中输入“$B$1：$B$9”，在“输出选项”中选择“新工作表组”，如图 8-40 所示。

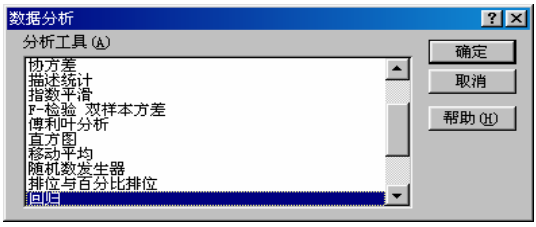


图 8-39　“数据分析”对话框

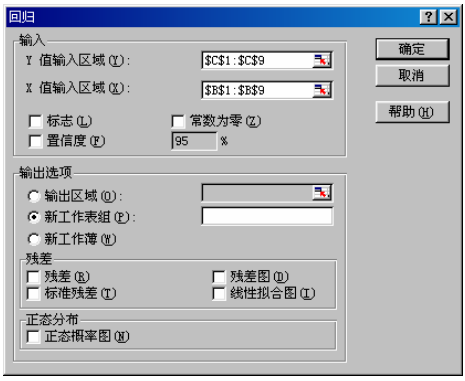


图 8-40　“回归”对话框

第三步：单击“确定”按钮，回归分析结果如图 8-41 所示。

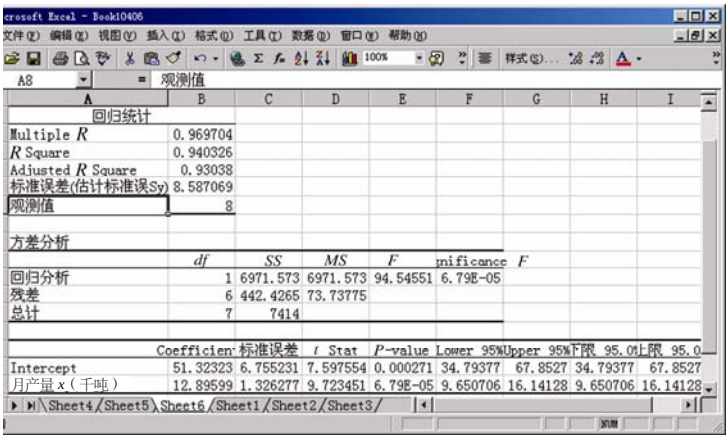


图 8-41　Excel 回归分析结果

在结果中，第一部分是回归统计；第二部分是方差分析；第三部分是回归系数表，其中 Intercept 指截距，即一元线性方程中的 a。根据上图可以求得，一元线性方程为

y=51.32+12.89x。

8.6 用Excel进行时间数列分析

8.6.1 计算时间数列分析指标

1．测定增长量和平均增长量

根据表 8-6 中的数据，计算逐期增长量、累积增长量和平均增长量。计算结果如图 8-42 所示。

表 8-6　国内生产总值

年份	国内生产总值 y
1995	58 478.1
1996	67 884.6
1997	74 462.6
1998	78 345.2
1999	82 067.5
2000	89 442.2
2001	97 314.8
2002	104 790.6
2003	116 694.0
2004	136 515.0

计算步骤如下。

第一步：在 A 列输入“年份”，在 B 列输入“国内生产总值”。

第二步：计算逐期增长量：在 C3 单元格输入公式“=B3–B2”，并用鼠标拖动填充句柄将公式复制到 C4：C11 区域。

第三步：计算累积增长量：在 D3 单元格输入公式“=B3–$B$2”，并用鼠标拖动填充句柄将公式复制到 D4：D11 区域。

第四步：计算平均增长量：在 C13 中输入公式“=（B11–B2）/8”，按回车键，即可得到平均增长量。

2．测定发展速度和平均发展速度

仍以表 8-6 中的国内生产总值为例，说明如何计算定基发展速度、环比发展速度和平均发展速度。计算结果如图 8-43 所示。

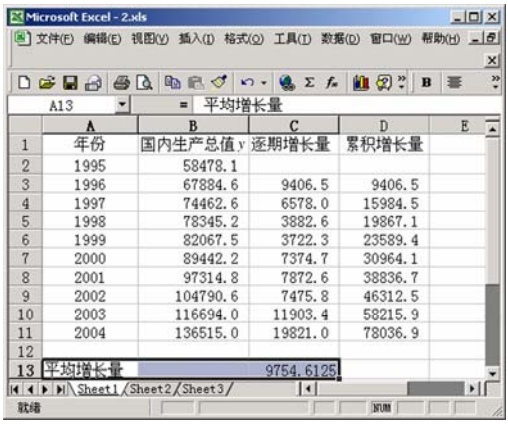


图 8-42　用 Excel 计算增长量和平均增长量

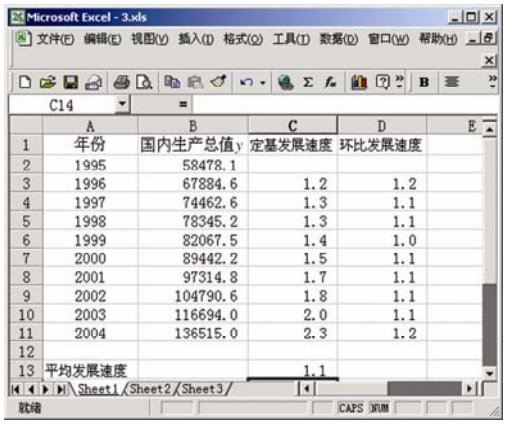


图 8-43　用 Excel 计算发展速度和平均发展速度

计算步骤如下。

第一步：在 A 列输入“年份”，在 B 列输入“国内生产总值”。

第二步：计算定基发展速度：在 C3 单元格输入公式“=B3/$B$2”，并用鼠标拖动填充句柄将公式复制到 C4：C11 区域。

第三步：计算环比发展速度：在 D3 单元格输入公式“= B3/B2”，并用鼠标拖动填充句柄将公式复制到 D4：D11 区域。

第四步：计算平均发展速度：在 C13 单元格中输入公式“=GEOMEAN（D3：D11）”，按回车键，即可得到平均增长量。或者选中 C13 单元格，单击“插入”菜单，选择“函数”选项，在如图 8-44 所示的对话框中选择 GEOMEAN（返回集合平均值）函数，在数值区域中输入“D3：D11”即可。

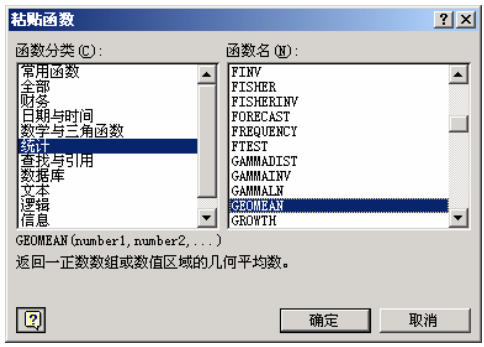


图 8-44　“粘贴函数”对话框

8.6.2　测定长期趋势

1．移动平均法

利用表 8-6 中的数据来说明如何用移动平均法计算长期趋势，计算结果如图 8-45 所示。

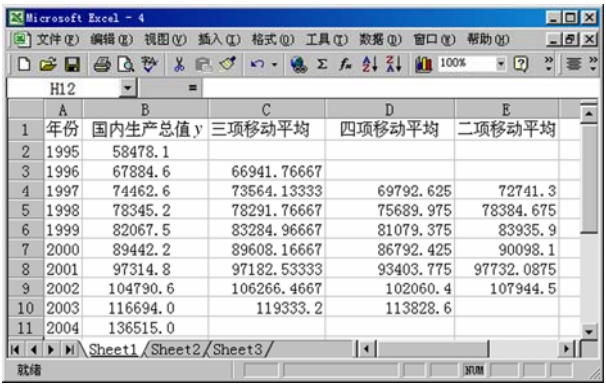
计算步骤如下。

第一步：在 A 列输入“年份”，在 B 列输入“国内生产总值”。

第二步：计算三项移动平均：在 C3 单元格中输入“=（B2+B3+B4）/3”，并用鼠标拖动填充句柄将公式复制到 C4：C10 区域。

第三步：计算四项移动平均：在 D4 单元格中输入“=SUM（B2：B5）/4”，并用鼠标

拖动填充句柄将公式复制到 D5：D10 区域。



	A	B	C	D	E
1	年份	国内生产总值 y	三项移动平均	四项移动平均	二项移动平均
2	1995	58478.1			
3	1996	67884.6	66941.76667		
4	1997	74462.6	73564.13333	69792.625	72741.3
5	1998	78345.2	78291.76667	75689.975	78384.675
6	1999	82067.5	83284.96667	81079.375	83935.9
7	2000	89442.2	89608.16667	86792.425	90098.1
8	2001	97314.8	97182.53333	93403.775	97732.0875
9	2002	104790.6	106266.4667	102060.4	107944.5
10	2003	116694.0	119333.2	113828.6	
11	2004	136515.0			

图 8-45　用移动平均法计算长期趋势

2．直线趋势方程

根据表 8-6 中国内生产总值的数据求直线趋势方程。

可以用如下方程组来求直线趋势方程 $y=a+bt$：

$$\begin{cases} \sum_{i=1}^{n} y = na + b\sum_{i=1}^{n} t \\ \sum_{i=1}^{n} ty = a\sum_{i=1}^{n} t + b\sum_{i=1}^{n} t^2 \end{cases}$$

求得直线趋势方程的系数是：

$$\begin{cases} a = \dfrac{\sum_{i=1}^{n} y - b\sum_{i=1}^{n} t}{n} \\ b = \dfrac{n\sum_{i=1}^{n} ty - \sum_{i=1}^{n} t\sum_{i=1}^{n} y}{n\sum_{i=1}^{n} t^2 - (\sum_{i=1}^{n} t)^2} \end{cases}$$

计算步骤如下。

第一步：将数据输入。

第二步：给时间值 t 分别赋予 1～10。

第三步：计算 t^2。在 D2 单元格输入公式“=B2*B2”。

第四步：计算 ty。在 E2 单元格输入公式“=B2*C2”。

第五步：计算直线趋势方程的系数。

先计算 b，在 B15 单元格中输入公式“=（10*E12–B12*C12）/（10*D12–B12*B12）”。再计算 a，在 B14 单元格中输入公式“=（C12–B15*B12）/10”。即所求直线趋势方程为 y = 48602.08 + 7635.88 t，如图 8-46 所示。

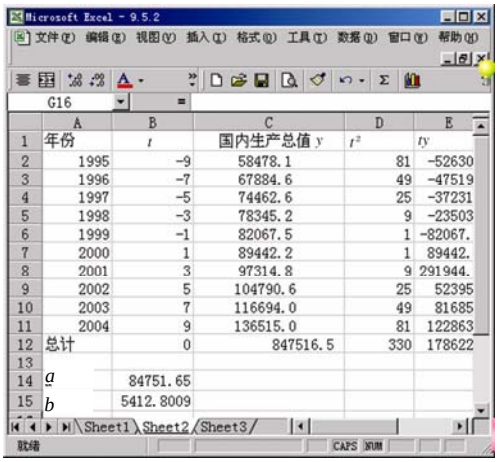
如果在设置 t 时，能使 $\sum t = 0$ 将大大简化计算。在上面的例子中，如果将 t 设置为–9，–7，–5，–3，–1，1，3，5，7，9，步骤 1～4 与上题一样。在计算方程系数时：先计算 a，在 B14 单元格中输入公式“= C12/10”。再计算 b，在 B15 单元格中输入公式“=E12/D12”，即所求方程为 y = 84751.65 + 5412.8 t。

计算结果如图 8-47 所示。



	A	B	C	D	E
1	年份	时间值 t	国内生产总值 y	t^2	ty
2	1995	1	58478.1	1	58478.1
3	1996	2	67884.6	4	135769.2
4	1997	3	74462.6	9	223387.8
5	1998	4	78345.2	16	313380.8
6	1999	5	82067.5	25	410337.5
7	2000	6	89442.2	36	536653.2
8	2001	7	97314.8	49	681203.6
9	2002	8	104790.6	64	838324.8
10	2003	9	116694.0	81	1050246
11	2004	10	136515.0	100	1365150
12	总计：	55	905994.6	385	5612931
13					
14	a	48602.08			
15	b	7635.887			

图 8-46　直线趋势方程计算结果



	A	B	C	D	E
1	年份	t	国内生产总值 y	t^2	ty
2	1995	-9	58478.1	81	-52630
3	1996	-7	67884.6	49	-47519
4	1997	-5	74462.6	25	-37231
5	1998	-3	78345.2	9	-23503
6	1999	-1	82067.5	1	-82067.
7	2000	1	89442.2	1	89442.
8	2001	3	97314.8	9	291944.
9	2002	5	104790.6	25	52395
10	2003	7	116694.0	49	81685
11	2004	9	136515.0	81	122863
12	总计	0	847516.5	330	178622
13					
14	a	84751.65			
15	b	5412.8009			

图 8-47　直线趋势方程计算结果

8.6.3　季节变动

表 8-7 是某服装公司销售额资料。

表 8-7　某服装公司销售额资料

月　份	各年销售额				
	1996	1997	1998	1999	2000
	（1）	（2）	（3）	（4）	（5）
1	1.1	1.1	1.4	1.4	1.3
2	1.2	1.5	2.1	2.1	2.2
3	1.9	2.2	3.1	3.1	3.3
4	3.6	3.9	5.2	5.0	4.9

续表

月　份	各年销售额				
	1996	1997	1998	1999	2000
	（1）	（2）	（3）	（4）	（5）
5	4.2	6.4	6.8	6.6	7.0
6	4.2	16.4	18.8	19.5	20.0
7	24.0	28.0	31.0	31.5	31.8
8	9.5	12.0	14.0	14.5	15.3
9	3.8	3.9	4.8	4.9	5.1
10	1.8	1.8	2.4	2.5	2.6
11	1.2	1.3	1.2	1.4	1.4
12	0.9	1.0	1.1	1.2	1.1

1．按月平均法

当时间数列的长期趋势不存在或不明显时，可采用按月（季）平均法。

利用表8-7中的数据计算季节比率。

计算步骤如下。

第一步：将数据输入Excel中。

第二步：求出每年的年总计。在B16单元格中输入公式“=SUM（B4：B15）”，并将公式复制到C16：F16。

第三步：求出每年同月销售额总计。在G4单元格中输入公式“=SUM（B4：F4）”，并将公式复制到G5：G16。

第四步：求出每年同月销售额平均值。在H4单元格中输入公式“=AVERAGE（B4：F4）”，并将公式复制到H5：H15。

第五步：求得总的月销售额平均数。在H16单元格中输入公式“=AVERAGE（B4：F15）”或“=AVERAGE（H4：H15）”或“=G16/60”。

第六步：求季节比率。在I4单元格中输入公式“=H4/$H$16”，并将公式复制到I5：I15。

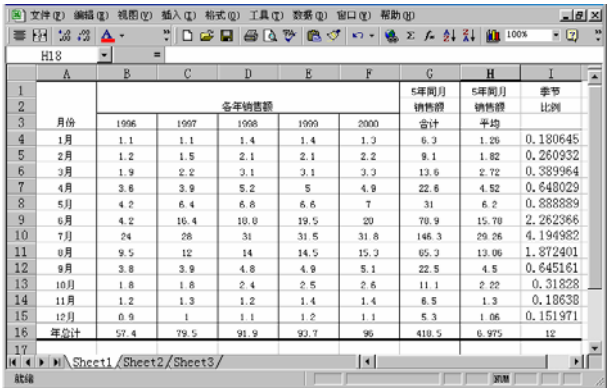
第七步：在I16单元格中输入公式“=SUM（I4：I15）”。

计算出季节比率之和为12，无须调整，最后结果如图8-48所示。

2．趋势剔除法

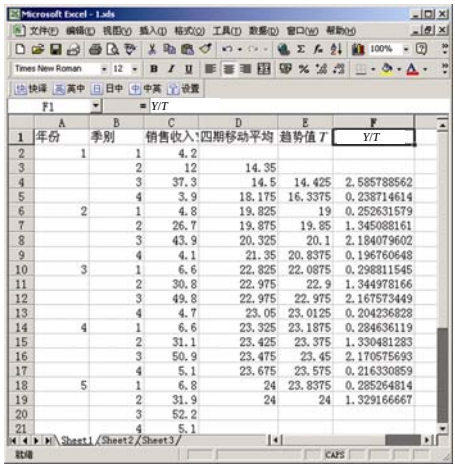
当时间数列存在明显的长期趋势时，需要先剔除长期趋势的影响，然后再计算季节比率。

利用表8-7中的数据，说明如何用移动平均趋势剔除法测定季节变动，如图8-49所示。



	A	B	C	D	E	F	G	H	I
1							5年周月	5年周月	季节
2		各年销售额					销售额	销售额	比例
3	月份	1996	1997	1998	1999	2000	合计	平均	
4	1月	1.1	1.1	1.4	1.4	1.3	6.3	1.26	0.180645
5	2月	1.2	1.5	2.1	2.1	2.2	9.1	1.82	0.260932
6	3月	1.9	2.2	3.1	3.1	3.3	13.6	2.72	0.389964
7	4月	3.8	3.9	5.2	5	4.9	22.8	4.52	0.648029
8	5月	4.2	6.4	6.8	6.6	7	31	6.2	0.888889
9	6月	4.2	16.4	10.8	19.5	20	70.9	15.78	2.262366
10	7月	24	28	31	31.5	31.8	146.3	29.26	4.194982
11	8月	9.5	12	14	14.5	15.3	65.3	13.06	1.872401
12	9月	3.8	3.9	4.8	4.9	5.1	22.5	4.5	0.645161
13	10月	1.8	1.8	2.4	2.5	2.6	11.1	2.22	0.31828
14	11月	1.2	1.3	1.2	1.4	1.4	6.5	1.3	0.18638
15	12月	0.9	1	1.1	1.2	1.1	5.3	1.06	0.151971
16	年合计	57.4	79.5	91.9	93.7	96	418.5	6.975	12
17									

图 8-48　按月平均法计算长期趋势



	A	B	C	D	E	F
1	年份	季别	销售收入	四期移动平均	趋势值 T	Y/T
2	1	1	4.2			
3		2	12	14.35		
4		3	37.3	14.5	14.425	2.585788562
5		4	3.9	18.175	16.3375	0.238714614
6	2	1	4.8	19.825	19	0.252631579
7		2	26.7	19.875	19.85	1.345088161
8		3	43.9	20.325	20.1	2.184079602
9		4	4.1	21.35	20.8375	0.196760648
10	3	1	6.6	22.825	22.0875	0.298811545
11		2	30.8	22.975	22.9	1.344978166
12		3	49.8	22.975	22.975	2.167573449
13		4	4.7	23.05	23.0125	0.204236828
14	4	1	6.6	23.325	23.1875	0.284636119
15		2	31.1	23.425	23.375	1.330481283
16		3	50.9	23.475	23.45	2.170575693
17		4	5.1	23.675	23.575	0.216330859
18	5	1	6.8	24	23.8375	0.285264814
19		2	31.9	24	24	1.329166667
20		3	52.2			
21		4	5.1			

图 8-49　用剔除趋势法测定季节变动

计算步骤如下。

第一步：按下图的格式在 A 列输入“年份”，在 B 列输入“季别”，在 C 列输入“销售收入”。

第二步：计算四期移动平均：在 D3 中输入“=SUM（C2：C5）/4”，并用鼠标拖拽将公式复制到 D4：D19 区域。

第三步：计算趋势值（即二项移动平均）T：在 E4 单元格中输入“=（D3+D4）/2”，并拖拽鼠标将公式复制到 E5：E19 区域。

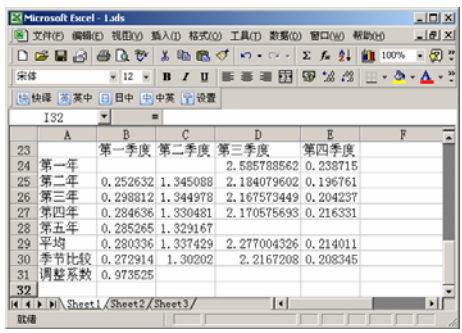
第四步：剔除长期趋势，即计算 Y/T：在 F4 单元格中输入“=C4/E4”，并拖拽鼠标将公式复制到 F5：F19 区域。

第五步：重新排列 F5：F19 中的数据，使同季的数字位于一列，共排成四列。

第六步：计算各年同季平均数：在 B29 单元格中输入公式“=AVERAGE（B24：B28）”，在 C29 中输入公式“=AVERAGE（C24：C28）”，在 D29 中输入公式“=AVERAGE（D24：D28）”，在 E29 中输入公式“=AVERAGE（E24：E28）”。

第七步：计算调整系数：在 B31 单元格中输入公式“=4/SUM（B29：E29）”。

第八步：计算季节比率：在 B30 中输入公式“=B29*$B$31”，并拖拽鼠标将公式复制到 B30：E30 区域，就可以得到季节比率的值，如图 8-50 所示。



	A	B	C	D	E	F
23		第一季度	第二季度	第三季度	第四季度	
24	第一年			2.585788562	0.238715	
25	第二年	0.252632	1.345088	2.184079602	0.196761	
26	第三年	0.298812	1.344978	2.167573449	0.204237	
27	第四年	0.284636	1.330481	2.170575693	0.216331	
28	第五年	0.285265	1.329167			
29	平均	0.280336	1.337429	2.277004326	0.214011	
30	季节比较	0.272914	1.30202	2.2167208	0.208345	
31	调整系数	0.973525				
32						

图 8-50　用 Excel 计算季节比率

8.7 用Excel进行指数分析

8.7.1 计算总指数

指数分析法是研究社会经济现象数量变动情况的一种统计分析方法。指数有综合指数和平均指数之分，本节介绍如何用 Excel 进行指数分析。

1．综合指数

图 8-51 中是某企业三种产品的生产情况，以基期价格 p_0 作为同度量因素，计算生产指数，结果如图 8-51 所示。

计算步骤如下。

第一步：计算各个 p_0q_0，在 G2 中输入“=E2*C2”，并拖拽鼠标将公式复制到 G3：G4 区域。

第二步：计算各个 p_0q_1：在 H2 中输入“=E2*D2”，并拖拽鼠标将公式复制到 H3：H4 区域。

第三步：计算 $\sum p_0q_0$ 和 $\sum p_0q_1$：选定 G2：G4 区域，单击工具栏上的 $\sum$ 按钮，在 H5 单元格中将出现该列的求和值。

第四步：计算生产量综合指数 $Iq=\dfrac{p_0q_1}{p_0q_0}$：在 C6 中输入“=H5/G5”便可得到生产量综合指数。

2．平均数指数

（1）加权平均数指数

已知某企业有关生产情况，计算平均数指数，计算结果如图 8-52 所示。

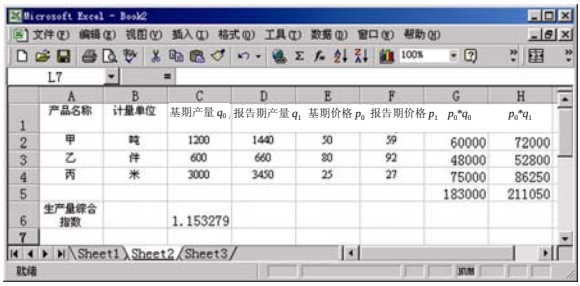


图 8-51　用 Excel 计算总指数结果



图 8-52　用 Excel 计算加权平均数

计算步骤如下。

第一步：在 D2 单元格中输入公式“=B2*C2/100”。

第二步：在 D5 单元格中输入公式“=SUM（D2：D4）”。

第三步：在 B7 单元格中输入公式“=D5/B5”，即得到所求的值。

（2）加权调和平均数指数

已知某企业三种产品的生产情况，求产品价格总的变化情况，求得的结果如图 8-53 所示。

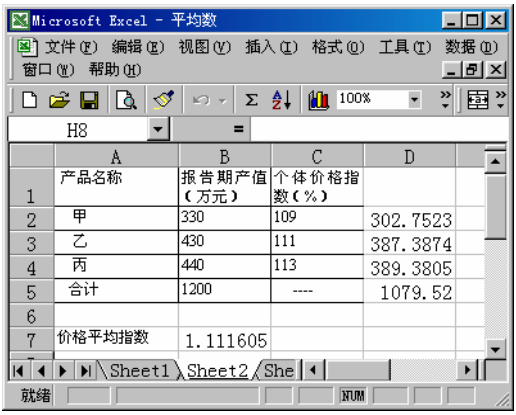


图 8-53　用 Excel 计算加权调和平均数

计算步骤如下。

第一步：在 D2 单元格中输入公式“=B2/C2*100”。

第二步：在 D5 单元格中输入公式“=SUM（D2：D4）”。

第三步：在 B7 单元格中输入公式“=B5/D5”即得到所求的值。

8.7.2　因素分析

利用表 8-8 中的数据进行因素分析。

表 8-8　某商场三种商品销售资料

商品名称	计量单位	销售量 q		价格 p（元）	
		基期 q_0	报告期 q_1	基期 p_0	报告期 p_1
甲	套	300	380	280	260
乙	件	450	510	140	150
丙	米	1 200	1 500	70	78

计算步骤如下。

第一步：计算各个 p_0*q_0 ：在 G3 单元格中输入公式“=E3*C3”，并将公式复制到 G4：G5。

第二步：计算各个 p_0*q_1：在 H3 单元格中输入公式“=E3*D3”，并将公式复制到 H4：H5。

第三步：计算各个 p_1*q_1：在 I3 单元格中输入公式“=F3*D3”，并将公式复制到 I4：I5。

第四步：计算$\sum p_0*q_0$，$\sum p_0*q_1$，$\sum p_1*q_1$：在 G6 单元格中单击工具栏上的$\sum$按钮，并将公式复制到 H6：I6。

第五步：计算总成本指数：在 B7 单元格中输入“=I6/G6”，即求得总成本指数。

第六步：计算产量指数：在 B8 单元格中输入“=H6/G6”，即求得产量指数。

第七步：计算单位成本指数：在 B9 单元格中输入“=I6/H6”，即求得单位成本指数。

最后结果如图 8-54 所示。



	A	B	C	D	E	F	G	H	I
1	商品	计量	销售量 q		价格 p（元）		p_0*q_0	p_0*q_1	p_1*q_1
2	名称	单位	基期 q_0	报告期 q_1	基期 p_0	报告期 p_1			
3	甲	套	300	380	280	260	84000	106400	98800
4	乙	件	450	510	140	150	63000	71400	76500
5	丙	米	1200	1500	70	78	84000	105000	117000
6							231000	282800	292300
7	总成本指数	1.265368							
8	产量指数	1.224242							
9	单位成本指数	1.033593							

图 8-54　用 Excel 进行因素分析结果

本章小结

1. 计算机软件在数据统计中发挥着重要的作用，越来越不可替代。

2. 在软件安装过程中一定要在 Excel 的“工具”栏目先选中“加载宏”，打开对话框选中“分析工具库”，然后安装，就可以实现基本数据统计以及数据分析。

3. 练习过程要多次联系，了解有关数据结果的实现和对结果的解释，特别是熟练运用 Excel 软件对回归模型参数的实现和解读，这些内容在实际工作中有非常广泛的应用。

复习思考题

1. 常州老九管理咨询有限公司在市场收集了咨询单位生产某种工业添加剂的市场价格和市场需求量的 7 组数据（见表 8-9），请利用计算机软件实现需要数据结果。

表 8-9　企业产品价格和销售量数据表

编号	价格 x（元/公斤）	需求量 y（公斤）
1	7	75
2	6	80
3	5	70
4	8	60
5	7	65

续表

编号	价格 x（元/公斤）	需求量 y（公斤）
6	5	85
7	4	90
合计		

问题：1）计算价格和产品需求量之间的线性相关系数。

2）拟合价格和需求量的直线回归方程，并且说明回归系数的实际意义。

3）计算估计标准误差，说明回归方程的拟合程度。

4）计算判定系数，说明回归方程的拟合程度。

5）以 5%的显著性水平估计价格为 10 元时需求量值。

2. 某地区棉花产量的年度资料如表 8-10 所示，请用最小平方法编制一个直线趋势方程式，解释回归方程的可靠程度，并预计 2009 年棉花产量。

表 8-10　2002—2008 年某地区棉花产量

年份	产 量（吨）
2002	354.0
2003	424.5
2004	414.9
2005	378.8
2006	450.8
2007	567.5
2008	450.8

附录　附　表

附表 1　正态分布概率表

t	$F(t)$	t	$F(t)$	t	$F(t)$	t	$F(t)$
0.00	0.000 0	0.53	0.403 9	1.06	0.710 9	1.59	0.888 2
0.01	0.008 0	0.54	0.410 8	1.07	0.715 4	1.60	0.890 4
0.02	0.016 0	0.55	0.417 7	1.08	0.719 9	1.61	0.892 6
0.03	0.023 9	0.56	0.424 5	1.09	0.724 3	1.62	0.894 8
0.04	0.031 9	0.57	0.431 3	1.10	0.728 7	1.63	0.896 9
0.05	0.039 9	0.58	0.438 1	1.11	0.733 0	1.64	0.899 0
0.06	0.047 8	0.59	0.444 8	1.12	0.737 3	1.65	0.901 1
0.07	0.055 8	0.60	0.451 5	1.13	0.741 5	1.66	0.903 1
0.08	0.063 8	0.61	0.458 1	1.14	0.745 7	1.67	0.905 1
0.09	0.071 7	0.62	0.454 7	1.15	0.749 9	1.68	0.907 0
0.10	0.079 7	0.63	0.471 3	1.16	0.754 0	1.69	0.909 0
0.11	0.087 6	0.64	0.477 8	1.17	0.758 0	1.70	0.910 9
0.12	0.095 5	0.65	0.484 3	1.18	0.762 0	1.71	0.912 7
0.13	0.103 4	0.66	0.490 7	1.19	0.766 0	1.72	0.914 6
0.14	0.111 3	0.67	0.497 1	1.20	0.769 9	1.73	0.916 4
0.15	0.119 2	0.68	0.503 5	1.21	0.773 7	1.74	0.918 1
0.16	0.127 1	0.69	0.509 8	1.22	0.777 5	1.75	0.919 9
0.17	0.135 0	0.70	0.516 1	1.23	0.781 3	1.76	0.921 6
0.18	0.142 8	0.71	0.522 3	1.24	0.785 0	1.77	0.923 3
0.19	0.150 7	0.72	0.528 5	1.25	0.788 7	1.78	0.924 9
0.20	0.158 5	0.73	0.534 6	1.26	0.792 3	1.79	0.926 5
0.21	0.166 3	0.74	0.540 7	1.27	0.795 9	1.80	0.928 1
0.22	0.174 1	0.75	0.546 7	1.28	0.799 5	1.81	0.929 7
0.23	0.181 9	0.76	0.552 7	1.29	0.803 0	1.82	0.931 2
0.24	0.189 7	0.77	0.558 7	1.30	0.806 4	1.83	0.932 8

续表

t	$F(t)$	t	$F(t)$	t	$F(t)$	t	$F(t)$
0.25	0.197 4	0.78	0.564 6	1.31	0.809 8	1.84	0.934 2
0.26	0.205 1	0.79	0.570 5	1.32	0.813 2	1.85	0.935 7
0.27	0.212 8	0.80	0.576 3	1.33	0.816 5	1.86	0.937 1
0.28	0.220 5	0.81	0.582 1	1.34	0.819 8	1.87	0.938 5
0.29	0.228 2	0.82	0.587 8	1.35	0.823 0	1.88	0.939 9
0.30	0.235 8	0.83	0.593 5	1.36	0.826 2	1.89	0.941 2
0.31	0.243 4	0.84	0.599 1	1.37	0.829 3	1.90	0.942 6
0.32	0.251 0	0.85	0.604 7	1.38	0.832 4	1.91	0.943 9
0.33	0.258 6	0.86	0.610 2	1.39	0.835 5	1.92	0.945 1
0.34	0.266 1	0.87	0.615 7	1.40	0.838 5	1.93	0.946 4
0.35	0.273 7	0.88	0.621 1	1.41	0.841 5	1.94	0.947 6
0.36	0.281 2	0.89	0.626 5	1.42	0.844 4	1.95	0.948 8
0.37	0.288 6	0.90	0.631 9	1.43	0.847 3	1.96	0.950 0
0.38	0.296 1	0.91	0.637 2	1.44	0.850 1	1.97	0.951 7
0.39	0.303 5	0.92	0.642 4	1.45	0.852 9	1.98	0.952 3
0.40	0.310 8	0.93	0.647 6	1.46	0.855 7	1.99	0.953 4
0.41	0.318 2	0.94	0.652 8	1.47	0.858 4	2.00	0.954 5
0.42	0.325 5	0.95	0.657 9	1.48	0.861 1	2.10	0.964 3
0.43	0.332 8	0.96	0.662 9	1.49	0.863 8	2.20	0.972 2
0.44	0.340 1	0.97	0.668 0	1.50	0.866 4	2.30	0.978 6
0.45	0.347 3	0.98	0.672 9	1.51	0.869 0	2.40	0.983 6
0.46	0.354 5	0.99	0.677 8	1.52	0.871 5	2.50	0.987 6
0.47	0.361 6	1.00	0.682 7	1.53	0.874 0	2.60	0.990 7
0.48	0.368 8	1.01	0.687 5	1.54	0.876 4	2.70	0.993 1
0.49	0.375 9	1.02	0.692 3	1.55	0.878 9	2.80	0.994 9
0.50	0.382 9	1.03	0.697 0	1.56	0.881 2	2.90	0.996 2
0.51	0.389 9	1.04	0.701 7	1.57	0.883 6	3.00	0.997 3
0.52	0.396 9	1.05	0.706 3	1.58	0.885 9	4.00	0.999 4

附表 2　随机数字表

53	74	23	99	67	61	32	28	69	84	94	62	67	86	24	98	33	41	19	95	47	53	53	38	09
63	38	06	86	54	99	00	65	26	94	02	82	90	23	07	79	62	67	80	60	75	91	12	81	19
65	30	58	21	46	06	72	17	10	94	25	21	31	75	96	49	28	24	00	49	55	65	79	78	07
63	43	36	82	69	65	51	18	37	88	61	38	44	12	45	32	92	85	88	65	54	34	81	85	35
98	25	37	55	26	01	91	82	81	46	74	71	12	94	97	24	02	71	37	07	03	92	18	66	75
02	63	21	17	69	71	50	80	89	56	38	15	70	11	48	43	40	45	86	98	00	83	26	91	03
64	55	22	21	82	48	22	28	06	00	61	54	13	43	91	82	78	12	23	29	06	66	24	12	27
85	07	26	13	89	01	10	07	82	04	59	63	69	36	03	69	11	15	83	80	13	29	54	19	28
58	54	16	24	15	51	54	44	82	00	62	61	65	04	69	38	18	65	18	97	85	72	13	49	21
34	85	27	84	87	61	48	64	56	26	90	18	48	13	26	37	70	15	42	57	65	65	80	39	07
03	92	18	27	46	57	99	16	96	56	30	33	72	85	22	84	64	38	56	98	99	01	30	98	64
62	95	30	27	59	37	75	41	66	48	86	97	80	61	45	23	53	04	01	63	45	76	08	64	27
08	45	93	15	22	60	21	75	46	91	98	77	27	85	42	28	88	61	08	84	69	62	03	42	73
07	08	55	18	40	45	44	75	13	90	24	94	96	61	02	57	55	66	83	15	73	42	37	11	61
01	85	89	95	66	51	10	19	34	88	15	84	97	19	75	12	76	39	43	78	64	63	91	08	25
72	84	71	14	35	19	11	58	49	26	50	11	17	17	76	86	31	57	20	18	95	60	78	46	75
88	78	28	16	84	13	52	53	94	83	75	45	69	30	96	73	89	65	70	31	99	17	43	48	76
45	17	75	65	57	28	40	19	72	12	25	12	74	75	67	60	40	60	81	19	24	62	01	61	16
96	76	28	12	54	22	01	11	94	25	71	96	16	16	88	68	64	36	74	45	19	59	50	88	92
43	31	67	72	30	24	02	94	08	63	38	32	36	66	02	69	36	38	25	39	48	03	45	15	22
50	44	66	44	21	66	06	58	05	62	68	15	54	35	02	42	35	48	96	32	14	52	41	52	48
22	66	22	15	86	26	63	75	41	99	58	42	36	72	24	58	37	52	18	51	03	37	18	39	11
96	24	40	14	51	23	22	30	88	57	95	67	47	29	83	94	69	40	06	07	18	16	36	78	86
31	73	91	61	19	60	20	72	93	48	98	57	07	23	69	65	95	39	69	58	56	80	30	19	44
78	60	73	99	84	43	89	94	36	45	56	69	47	07	41	90	22	91	07	12	78	35	34	08	72
84	37	90	61	56	70	10	23	98	05	85	11	34	76	60	76	48	45	34	60	01	64	18	39	96
36	67	10	08	23	98	93	35	08	86	99	29	76	29	81	33	34	91	58	93	63	14	52	32	52
07	28	59	07	48	89	64	58	89	75	83	85	62	27	89	30	14	78	56	27	86	63	59	80	02
10	15	83	87	60	79	24	31	66	56	21	48	24	06	93	91	98	94	05	49	01	47	59	38	00
35	19	68	97	65	03	73	52	16	56	06	53	55	90	27	33	42	29	38	87	22	13	88	83	34
53	81	29	13	39	35	01	20	71	34	62	33	74	82	14	53	73	19	09	03	56	54	29	56	93
51	86	32	68	92	33	98	74	66	99	40	14	71	94	58	45	94	19	38	81	14	44	99	81	07

续表

35	91	70	29	13	80	03	54	07	27	96	94	78	32	66	50	95	52	74	33	13	80	55	62	54
37	71	67	95	16	20	02	44	95	94	64	85	04	05	72	01	32	90	76	14	53	89	74	60	41
93	66	13	83	27	92	79	64	64	72	28	54	96	53	84	48	14	52	98	94	56	07	93	89	30
02	96	08	45	65	13	05	00	41	84	93	07	54	72	59	21	45	57	09	77	19	48	56	27	44
49	83	43	48	35	82	88	33	69	96	72	36	04	19	76	47	45	15	18	60	82	11	08	95	97
84	60	71	62	46	40	80	81	30	37	34	39	23	05	38	25	15	35	71	30	88	12	57	21	77
18	17	30	88	71	44	91	14	88	47	89	23	30	63	15	56	34	20	47	89	99	82	93	24	98
79	69	10	61	78	71	32	76	95	62	87	00	22	58	40	92	54	01	75	25	43	11	71	99	31
75	93	36	57	83	56	20	14	82	11	74	21	97	90	65	96	42	68	63	86	74	54	13	26	94
38	30	92	29	03	06	28	81	39	38	62	25	06	84	63	61	29	08	93	67	04	32	92	08	09
51	28	50	10	34	31	57	75	95	80	51	97	02	74	77	76	15	48	49	44	18	55	63	77	09
21	31	38	86	24	37	79	81	53	74	73	24	16	10	33	52	83	90	94	76	70	47	14	54	36
29	01	23	87	88	58	02	39	37	67	42	10	14	20	92	16	55	23	42	45	54	96	09	11	06
95	33	95	22	00	18	74	72	00	18	38	79	58	69	32	81	76	80	26	92	82	80	84	25	39
90	84	60	79	80	24	36	59	87	38	82	07	53	89	35	96	35	23	79	18	05	98	90	07	35
46	40	62	98	82	54	97	20	56	95	15	74	80	08	32	16	46	70	80	80	67	72	16	42	79
20	31	89	03	43	38	46	82	68	72	32	14	82	99	70	80	60	47	18	97	66	39	67	98	60
71	59	73	05	50	08	22	23	71	77	91	01	93	20	49	82	96	59	26	94	66	39	67	98	60

参 考 文 献

[1] 黄良文. 统计学原理. 北京：中央广播电视大学出版社，2001
[2] 贾俊平等. 统计学. 北京：中国人民大学出版社，2000
[3] 栗方忠. 统计学原理. 大连：东北财经大学出版社，2001
[4] 邢西治等. 社会经济统计学原理. 南京：南京大学出版社，1998
[5] 骆克任，韩效宥. 现代实用统计与计算机应用. 上海：立信会计出版社，1996
[6] 陈珍珍等. 统计学. 厦门：厦门大学出版社，2002
[7] 郑聪玲. 统计. 杭州：浙江大学出版社，2004
[8] 凌洁. 统计学. 上海：上海财经大学出版社，2004
[9] 李卉妍等. 统计学. 北京：电子工业出版社，2007
[10] 孙允午. 统计学. 上海：上海财经大学出版社，2006
[11] 贾怀勤等. 应用统计. 北京：对外经贸大学出版社，2005
[12] 李洁明等. 统计学原理. 上海：复旦大学出版社，2006
[13] 曾五一. 统计学. 北京：中国金融出版社，2006
[14] 刘春英. 应用统计. 北京：清华大学出版社，2006
[15] [美]戴维·安德森，丹尼斯·斯威尼，托马斯·威廉斯. 商务与经济统计. 北京：机械工业出版社，2004